THE UNIVERSITY OF CHICAGO

MONETARY AND PSYCHOLOGICAL INCENTIVES ACROSS CULTURES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

DANILA MEDVEDEV

CHICAGO, ILLINOIS

AUGUST 2024

Table of Contents

List of Figures

List of Tables

List of Images

ACKNOWLEDGEMENTS

I am grateful to Thomas Talhelm for the trust we have developed and the freedom he has always given me. I am also grateful to the other members of my committee: Anuj Shah, whose feedback on this dissertation has been invaluable; Ayelet Fishbach, who is the reason I applied to the University of Chicago; and Josh Jackson, meeting whom toward the end of my PhD has been a gift. I will also always be grateful to Melissa Ferguson, my undergraduate advisor, one of the kindest and most intelligent people I have met. I repeat the acknowledgments, published elsewhere, to everyone who has participated in the data collected for this dissertation, in particular Yin Li.

Amira, Andrew, Angelina, Baird, Beatriz, Dima, Jaewuk, Masha, Nastya, Sam—whether we talk often, rarely, or not at all anymore, I feel love toward each of you and always will. I met many wonderful people—students, staff, and faculty—during my time at the University of Chicago. Diag Davenport—I am proud of you and incredibly grateful for the support and intellectual arguments we have shared since day one. Hannah Kim—you are smarter and cooler than me, and I love you unconditionally. Ben Morris—thank you for your kindness and patience with me, and for the beautiful energy you exude.

Above all, I thank my family. No amount of time or acculturation could fully make me grasp the series of events that transported me from the Mitino District of Moscow, Russia, to the University of Chicago in Chicago, Illinois, United States of America. What I do know is that none of this would have been possible without my mother. She came to Moscow from a mining town in Ukraine to study at a university and wanted me to do even more, even if it has brought her mostly pain. I love you, Mom.

ABSTRACT

Motivating effortful behavior is a problem that employers, governments, and nonprofits face globally. However, most studies on motivation are conducted in Western, educated, industrialized, rich, and democratic (WEIRD) cultures. We compared how hard crowd-source workers in a diverse set of nations worked in response to monetary incentives versus psychological motivators, such as competing with or helping others. In our first study, the advantage that money had over psychological interventions was larger in the United States and the United Kingdom than in China, India, Mexico, and South Africa. In our next study, we randomly assigned cultural frames through language in bilingual Facebook users in India. Money increased effort over a psychological treatment by 27% in Hindi and 52% in English. In the final study, we measured how likely crowd-source workers in 40 countries are to work beyond the minimum amount of time required to receive pay. Findings in this dissertation suggest that the market mentality of exchanging time and effort for material benefits is most prominent in the wealthy West, at least on low-wage work tasks.

Key Words: Culture, Incentives, Motivation, Psychological Contracts, Work

INTRODUCTION

"What do people get for all their hard work under the sun?" King Solomon asks at the beginning of Ecclesiastes (Bible, New Living Translation, 1996). In the modern world, work is instrumental to the goal of affording consumption. Through work, people receive money, which lets them pay their bills and, if they are lucky, one day buy a house.

Yet, motivation extends far beyond paychecks (Baron, 1988; Besley & Ghatak, 2008; Hansen et al., 2002; Herzberg et al., 1959; Mayo, 1933; Prelec & Bodner, 2003; Thibault Landry et al., 2017; Touré-Tillery & Fishbach, 2018; Weber, 1905). People strive to exceed personal expectations and make positive impacts in their communities (Hutson, 2000). They feel good when they work on tasks that align with their values (Deci et al., 2017; Deci & Ryan, 2000; Thibault Landry et al., 2017). For many, work provides meaning and a sense of identity (Gallus, 2017; Hogg & Terry, 2000; Karlsson et al., 2004; Rosso et al., 2010).

Both monetary and non-monetary incentives motivate people in and outside of the workplace (Ashraf et al., 2014; Cassar & Meier, 2018; Erkal et al., 2018; Frey & Gallus, 2017; Jenkins Jr et al., 1998; Touré-Tillery & Fishbach, 2018). But this knowledge alone may not satisfy employers who are searching for the most cost-effective ways to increase productivity among their workforce, or governments and nonprofits who must navigate tight budgets to fund public campaigns encouraging socially beneficial behaviors such as recycling or voting. Understanding the relative effectiveness of motivational strategies is essential to determine if the marginal benefit of implementing a monetary incentive justifies its higher cost.

Recent large-scale experiments directly contrasted the effectiveness of monetary and non-monetary incentives in motivating effort (Campos-Mercade et al., 2021; DellaVigna & Pope,

2018; Milkman et al., 2021). While both monetary and non-monetary incentives were found to be motivating, money had a large advantage over psychological motivators.

This line of work is important because it uses strong methods. The studies clearly compared the effectiveness of different incentives, with tight controls and random assignment. However, the samples in these experiments consisted almost entirely of people from the US and did not consider cultural differences. Can managers and policy-makers in India, Indonesia, or Nigeria apply scientific insights from these studies to issues facing their own countries?

Motivating Worldwide

Researchers have warned against indiscriminately exporting interventions that have demonstrated effectiveness in North America or Western Europe to other world regions (Brady et al., 2018; Domenech Rodríguez et al., 2018; Estrada-Villalta & Adams, 2018; Kizilcec & Cohen, 2017; Sinha, 1984; Sloan, 1990; C. C. Thomas & Markus, 2023). Yet to date, most interventions have rarely been tested outside of Western, educated, industrialized, rich, and democratic, "WEIRD," (Henrich et al., 2010) cultures (Burnette et al., 2022; Harguess et al., 2020; Liebe et al., 2018). For example, a 2018 meta-review looked at the effectiveness of "green nudges"— non-monetary incentives to use less electricity and switch to more environmentally friendly energy source. All 40 studies were conducted in high-income counties (World Bank, 2021), and 35 in the US or Western Europe.

This extreme sampling bias is not unique to intervention studies. Studies asking general questions about human nature face a similar problem (Henrich et al., 2010). Over 90% of all papers published in the top psychological journals use samples drawn exclusively from "WEIRD" cultures (Nielsen et al., 2017; Rad et al., 2018). Despite this sampling bias,

researchers then use these basic insights to develop interventions assuming that the findings are universal (Cheon et al., 2020; Rad et al., 2018).

Although rare, psychological interventions designed with a particular cultural context in mind have shown promise in addressing a range of pressing issues (Bandiera et al., 2017; Bossuroy et al., 2022; Bursztyn et al., 2019; Dalton et al., 2021; Kapoor et al., 2023; Wu & Paluck, 2021). For example, a study in China tried to encourage workers in a textile factory to throw their waste in trash cans instead of on the production floor (Wu & Paluck, 2021). The researchers found that placing printed images of golden coins (symbols of luck and good fortune in China) on the floor reduced littering, while fining people real money failed. Similarly, an intervention in Niger tried to encourage women to start their own businesses by paying them money or by changing community norms around female entrepreneurship (Bossuroy et al., 2022). The norms worked better per dollar spent than the cash transfers.

Psychological Contracts and Cultural Embeddedness of Economic Behavior

These studies highlight that economic activities are embedded within systems of local norms, prevalent morality, and social networks (Mauss, 2000; Polanyi, 1957a, 1957b; Scott, 1976; Thompson, 1971). Socio-cultural factors interact with formal institutions to shape economic judgments and behaviors, sometimes contradicting standard economic logic (Alesina & Giuliano, 2015; Boyer & Petersen, 2018; M. K. Chen, 2013; Fernández & Fogli, 2009; Fisman & Miguel, 2007; Greif, 1993; Guiso et al., 2006; Henrich et al., 2005; Markus, 2016; Roth et al., 1991; Shao et al., 2010).

A classic example of the clash between culture and an economic logic comes from E.P. Thompson's analysis of the 1776 food riots in England (Thompson, 1971). Historically, the Assize of Bread and Ale controlled the pricing of bread so that the loaves could reduce in size

but not rise in price when wheat was scarce. After several poor harvests led to severe shortages, merchants eventually raised the price of bread, arguing that further reducing the loaf size was economically unfeasible. The peasants did not accept that price increases were just and responded with violent protests. They saw the hikes not as a natural application of the laws of supply and demand but as a betrayal of the social contract and a direct threat to the moral fabric of their societies.

In work contexts, cultural embeddedness is crucial for understanding the nature of obligations that exist between employers and their employees (Adler & Jelinek, 1986; Biernacki, 1995; Blau & Scott, 1962), given that employment is both an economic and a social relationship (Baron & Kreps, 2012). Labor participation in the market economies of "WEIRD" cultures is perceived to occur in the economic domain, differentiated from the social or moral spheres (Durkheim, 1984; Harwick, 2023; Simmel, 1907). As such, it is said to be enacted in an impersonal and transactional manner, in accordance with the norms of market exchange (Fiske, 1991; Li & Zahra, 2012; Tiessen, 1997; Westwood et al., 2004). Employees must work a specified amount of time or produce a specified amount of output to receive a specified amount of money from the employer. The scope of the mutual obligations is explicitly outlined in a formal contract (Raja et al., 2004; D. C. Thomas et al., 2003a).

Outside of "WEIRD" cultures, work usually operates under norms of reciprocity exchange, which are marked by "contractual incompleteness" (Bowles, 1998; Westwood et al., 2004). In such "relational contracts" (Raja et al., 2004; Rousseau & McLean Parks, 1993), people usually do not rely on the strict meanings of sentences specified in written agreements. Instead, they interpret obligations based on existing relationships and norms (Boyacigiller & Adler, 1991). Even when there is a formal agreement, people express some expectations tacitly,

and obligations often go beyond the terms specified in the contract (Rousseau & McLean Parks, 1993).

Of course, not all employment relations are governed solely by norms of market exchange in "WEIRD" cultures and norms of reciprocity exchange in "non-WEIRD" cultures. People from all cultures can apply different mental models of exchange norms and psychological contracts based on the features of a situation, including in economic transactions (Bandiera et al., 2010; DiMaggio, 1997; Gallus et al., 2022; Hui et al., 2004; Macchiavello, 2022; Mandel, 2006; Meeker, 1971). Nevertheless, cultural and institutional influences can make people more likely to interpret a situation as either transactional or relational by increasing the salience of a mental frame that is "consistent with what is likely or what ought to happen between employee and employer" (D. C. Thomas et al., 2003a, p. 458).

We argue that different exchange norms and psychological contracts influence the marginal advantage of monetary relative to non-monetary incentives, which we call the "money advantage." Workers from "WEIRD" cultures, operating under market exchange norms, should prioritize explicit *quid pro quo* arrangements—establishing what needs to be done in order to receive a tangible benefit (Rodríguez & Wilson, 2002; D. C. Thomas et al., 2003b)—and feel less obligated to reciprocate beyond what is contractually prescribed and formally enforced (Barron & Chou, 2016; Johnson & Droege, 2004).

In contrast, workers from "non-WEIRD" cultures, guided by reciprocity exchange norms, might not stop at the contractual minimum, even if they receive no extra money for doing so. They might intuit that this minimum does not represent their employer's real expectation (Gardner et al., 2015). Or they might reciprocate above the minimum prescribed by the contract to establish trust and earn loyalty for future interactions (Akerlof, 1982; Baron, 2013; Ghani &

Reed, 2022; Rodríguez & Wilson, 2002). Therefore, monetary incentives might have a smaller marginal effect on workers using reciprocity exchange norms compared to market exchange norms.

Previous Evidence on Work Incentives Across Cultures

Survey evidence has consistently shown that people from Western cultures report valuing money less as a source of job motivation (Chiang & Birtch, 2012; Corney & Richards, 2005; Ipeirotis, 2010; King & Bu, 2005). For instance, responding to a questionnaire about the desirability of different job characteristics, new IT recruits from the US reported prioritizing monetary bonuses for completed projects less than did their Chinese counterparts (King & Bu, 2005). People in the West are also more likely to say that work is overall less important to them. In the last wave of the World Values Survey, 92% of people in Egypt and 82% of people in China agreed that "work should always come first, even if it means less spare time." In the US and the UK, those percentages were 28% and 22%, respectively.

However, these studies measure attitudes and values, which predict behavior in some domains but not in others, often depending on whether they correspond to prevalent cultural norms and available schemas (House et al., 2004; Miles, 2015; Morris et al., 2015; Oyserman, 2015). People in the US and other "WEIRD" cultures tend to overemphasize the importance of extrinsic rewards, such as money, in their lay theories of what motivates others (Chatman & Barsade, 1995; Heath, 1999; Kasser et al., 2007; Miller, 1999; Vuolevi & Van Lange, 2010). In one study, researchers asked Citibank managers around the world what they thought was the main motivation of their employees (DeVoe & Iyengar, 2004). Compared to managers in Asia and Latin America, managers in the US were more likely to think that money and other extrinsic motivators were what drove their employees. Another study found that managers who choose to

motivate their employees through contingent rewards for performance with clear expectations for how to receive them are more positively evaluated in the West (Walumbwa et al., 2007).

Similarly, the few experiments that did directly compare the effectiveness of incentives across countries seem to suggest that money advantage is higher in "WEIRD" cultures (Bandiera et al., 2020; Davies & Fafchamps, 2022; Fryer Jr, 2011; Gneezy et al., 2019; Gorodnichenko & Roland, 2011; Jung & Avolio, 1999). For instance, in an experiment where the researchers tried paying students in the US and China to get better test scores, money improved performance in the US, but not in China (Gneezy et al., 2019).

Present Research

In the first four studies (Medvedev et al., 2024), we systematically compared the effectiveness of money in the US and the UK (two "WEIRD" cultures) versus China, India, Mexico, and South Africa (four "non-WEIRD" cultures) on motivating low-wage workers. In our experiments, workers received a fixed salary, a fixed salary plus a psychological intervention, or a fixed salary plus an additional monetary incentive. We tested whether money is equally effective in "WEIRD" and "non-WEIRD" cultures. We measured effectiveness as effort and as cost-effectiveness of different incentives. We then randomly assigned people in India to take the study in English or Hindi to see whether this shifted people's use of exchange norms and the motivational advantage of money. Finally, in the last study from an ongoing project, we measured effort in the absence of explicit monetary incentives in 40 countries and examined how moral norms and psychological contracts might shape whether people go above and beyond the contractually established minimum amount of effort.

STUDY 1: RE-ANALYSIS OF A LARGE-SCALE EXPERIMENT

In our first study, we re-analyzed data from a study by DellaVigna and Pope (DellaVigna & Pope, 2018), who gave Amazon Mechanical Turk (MTurk) workers a variety of incentives to perform a simple task. The task required participants to interchangeably click buttons "a" and "b" on their keyboards as many times as possible for 10 minutes. The researchers tested how much different incentives would increase the number of button-presses.

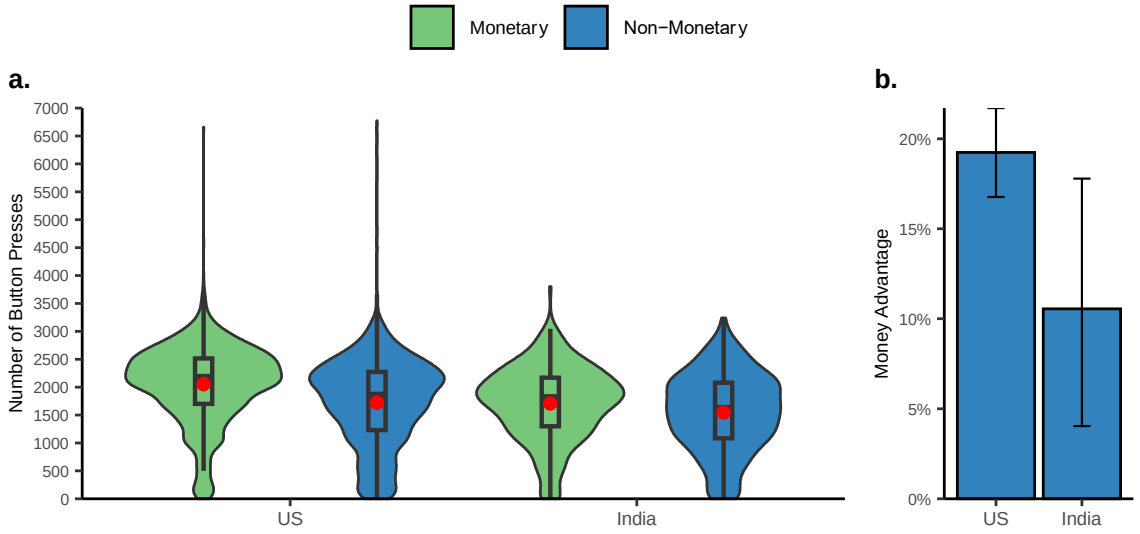
All participants received a $1.00 fee for participating in the experiment. On top of that, some people received monetary incentives, such as an extra cent for every 100 clicks. The researchers compared money to many incentives from social psychology. For example, in one condition, the researchers gave participants a social reference point: "most participants perform well on the task, pressing over 2,000 times." The results showed that (1) both pay-for-effort (monetary) and non-monetary incentives improved effort over the flat fee condition, which included no additional incentives, and (2) money outperformed non-monetary interventions by a large margin.

Fortunately for our studies, many of the recruited MTurk workers were in India. This gave us a convenient opportunity to test cultural differences—namely, whether the money advantage was higher in the US than in India. We excluded conditions with both a monetary and a psychological component (e.g., loss aversion; more information is available in the Methods section and Appendix Tables S27a and S27b). This left us with a sample of 5,526 participants from the US (3,196 female, $Mean_{\text{AgeCategory}} = 2.62$, $SD_{\text{AgeCategory}} = 1.29$) and 768 participants in India (247 female, $Mean_{\text{AgeCategory}} = 2.41$, $SD_{\text{AgeCategory}} = 1.02$). Using this combined sample, we compared the effectiveness of pooled monetary incentives versus pooled psychological interventions.

Results

The difference in effort between monetary incentives and psychological interventions was larger in the US than in India, as evidenced by an interaction between country and incentive type, $b = 170.56$, $t(6287) = 2.92$, $P = 0.004$, 95% CI 55.91 to 285.21 (Fig. 1 and Appendix Table S1). In these analyses, we controlled for age, gender, and education. Main effects of country and incentive type for this and the following studies are reported in section "Main Effects for Reported Regressions in Studies 1–4" in the Appendix. Additional Fig. 1 shows effort in the individual incentive treatments by country.

Fig. 1: **Study 1, Pooled Monetary Versus Pooled Non-Monetary Conditions in the US and India.**



Effects of pooled monetary incentives (green) and pooled non-monetary treatments (blue) in the US (N = 5,526 participants on MTurk) and India (N = 768 participants on MTurk) from a prior study (DellaVigna & Pope, 2018). Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median; the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive type in a multiple linear regression model is statistically significant, b = 170.56, $t(6287) = 2.92$, P = 0.004, 95% CI 55.91 to 285.21. Panel B shows the money advantage—that is, how much more effective money is than the pooled non-monetary treatments in each country. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of button-presses in the pooled monetary versus non-monetary treatments.

The results of Study 1 suggest that the money advantage differs across cultures. Specifically, monetary incentives outperformed psychological interventions more in the US than in India.

However, differences in technology could explain the results. For example, MTurk participants from the two countries could have understood the instructions differently. Furthermore, more participants in India could have completed the study on their phones or had slower internet connections, which could have limited how much they could ramp up their button-pressing, particularly in response to the monetary incentives.

STUDIES 2A–C: THE EFFECTIVENESS OF INCENTIVES IN 5 CULTURES

In Study 2, we improved on the weaknesses of Study 1 by designing a new task. In our new task, participants saw images and assessed whether each image showed a building. Participants rated images one by one for a maximum of 10 minutes. In the monetary conditions, we paid participants more for rating more images. The bonus across Studies 2a–2c ranged from 5 cents to 9 cents for every 10 images. In the non-monetary conditions, participants received the same pay regardless of how many images they rated. We asked comprehension questions to ensure that participants understood whether they would receive extra pay for rating more images.

We explicitly informed participants that they could quit the task without losing their base pay after rating 10 images. This gave a clear contractual minimal.

We also wanted to change the explicitly meaningless button-pressing task in Study 1 to be more like real-world work, which serves a purpose (Ariely et al., 2008; Cassar & Meier, 2018). To this end, we told participants that we needed their help in training a machine-learning image-classification algorithm.

Since the non-monetary conditions had the same payout structure, in what follows, we first report the results comparing monetary versus pooled non-monetary conditions, including the fixed-salary (or flat fee) condition. Then we compare the monetary incentive with each non-monetary treatment individually. In all regressions in Studies 2a–c, we control for gender, age, and education. Regressions with additional controls, including Internet connection, are reported in section "Additional Controls in Studies 2A–4" of the Appendix.

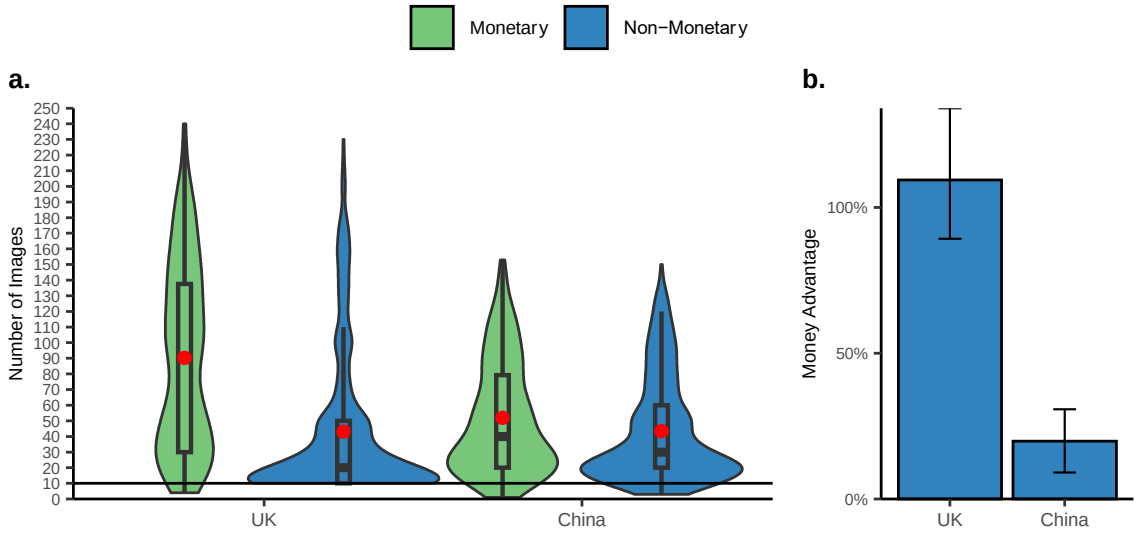
Study 2A: The UK and China

To test whether the results are generalizable to other cultures, we ran Study 2a in two new countries, the UK and China. To this end, we recruited 1,067 participants on Prolific in the UK (544 female, 12 non-binary, $Mean_{Age}$ = 40.04, SD_{Age} = 13.60) and 1,086 participants on social media through student networks at Hubei University in Wuhan, China (626 female, $Mean_{Age}$ = 23.31, SD_{Age} = 5.78). We compared the monetary condition (5 cents per 10 images) with the social norm and flat fee conditions.

Results

The difference in the effectiveness between monetary and non-monetary incentives was larger in the UK than in China, as evidenced by the interaction between incentive and culture, $b = 38.40$, $t(2145) = 9.65$, $P < 0.001$, 95% CI 30.59 to 46.20 (Fig. 2a and Appendix Table S2). In China, extra pay increased effort 19.9% over the two non-monetary conditions. In the UK, money increased effort 109.5% (Fig. 2b).

Fig. 2: **Study 2a, Monetary Versus Pooled Non-Monetary Conditions in the UK and China.**



Effects of a monetary incentive (green) and pooled non-monetary treatments (flat fee and social norm; blue) in the UK (N = 1,067 participants recruited on Prolific) and China (N = 1,086 participants recruited on social media). Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive type in a multiple linear regression model is statistically significant, $b = 38.40$, $t(2145) = 9.65$, $P < 0.001$, 95% CI 30.59 to 46.20. Panel B shows the money advantage—that is, how much more effective money is than the pooled non-monetary treatments in each country. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the monetary versus pooled non-monetary conditions.

The money advantage over each of the two non-monetary conditions (flat fee and norm) was significantly larger in the UK than in China, b = -44.88, t(2143) = -9.82, P < 0.001, 95% CI -53.84 to -35.91 for the flat fee condition, and b = -31.55, t(2143) = -6.96, P < 0.001, 95% CI -40.44 to -22.67 for the norm condition for the norm condition (Additional Fig. 2a and Appendix Table S7).

In China, the monetary condition was significantly less cost-effective than the norm condition, two-sided Welch's t(681.11) = -2.54, P = 0.011, P_{Bonf} = 0.045, $Mean_{\text{difference}}$ = -4.28, d = -0.19, 95 CI% -7.58 to -0.97, and did not significantly differ in cost-effectiveness from the flat fee condition, two-sided Welch's t(658.54) = 0.45, P = 0.653, P_{Bonf} = 1.000, $Mean_{\text{difference}}$ = 0.73, d = 0.03, 95 CI% -2.45 to 3.90 (Additional Fig. 2c). In the UK, the monetary condition was significantly more cost-effective than each of the two non-monetary conditions: Monetary and Norm, two-sided Welch's t(574.32) = 2.70, P = 0.007, P_{Bonf} = 0.029, $Mean_{\text{difference}}$ = 6.84, d = 0.20, 95 CI% 1.86 to 11.82; Monetary and Flat Fee, two-sided Welch's t(692.87) = 10.88, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 22.11, d = 0.81, 95 CI% 18.12 to 26.10.

Study 2B: Adjusting the Pay to Reflect the Real Value of Money in Mexico and the US

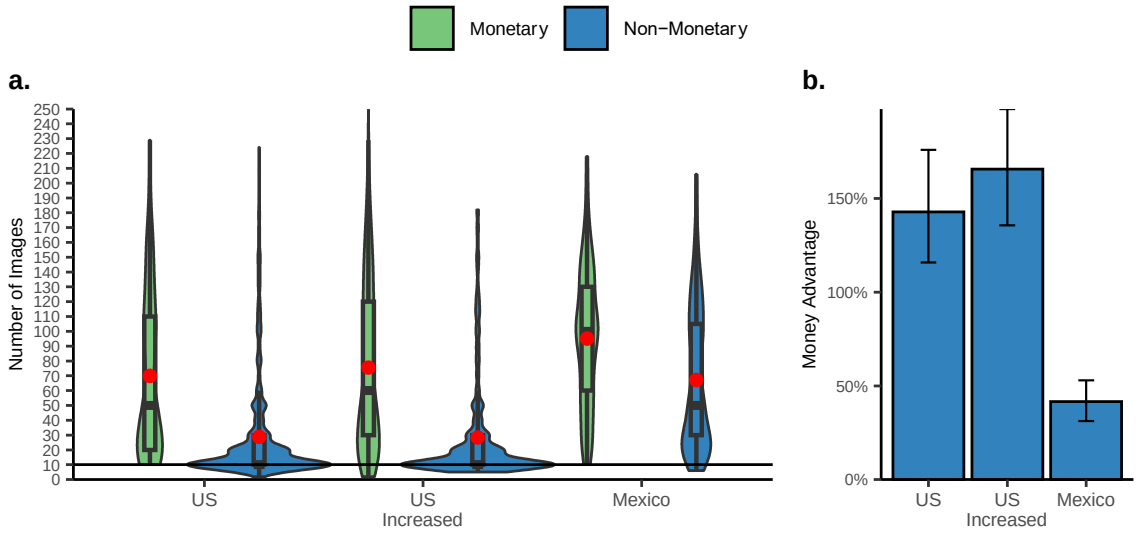
We followed up on Study 2a to make up for two shortcomings of the design. For one, Study 2a compared people on two different platforms (Prolific in the UK and social media in China). Because Prolific is a work platform, it could explain the difference in the money advantage without reference to culture. Second, we paid participants in the two countries the same amount of money. However, a dollar goes further in China than in the UK. To compensate for these shortcomings, in Studies 2b and 2c, we recruited workers only on crowdsourcing platforms and adjusted for local purchasing power.

In Study 2b, we extended the sample of "non-WEIRD" cultures to Mexico and again compared the monetary condition with the social norm and flat fee conditions. We recruited a Prolific sample in Mexico (N = 1,053; 536 female, 26 non-binary, $Mean_{Age}$ = 24.51, SD_{Age} = 5.43) and two Prolific samples in the US. In one of these US samples (N = 1,098; 652 female, 24 non-binary, $Mean_{Age}$ = 37.28, SD_{Age} = 13.77), we paid participants the same amount of money as in Mexico. In the other US sample (N = 1,122; 674 female, 25 non-binary, $Mean_{Age}$ = 36.13, SD_{Age} = 13.31), we increased the pay so that it became subjectively equivalent to the amount received by the participants in Mexico. To this end, we used a common economics procedure to establish subjective pay equivalence (Becker et al., 1964) (see the Methods section). Participants in the increased-pay sample in the US received $1.56 as base pay for their participation (compared to $1.30 in the other two samples) and a piece-rate of 6 cents (compared to 5 cents).

Results

The money advantage was larger in the US than in Mexico, regardless of whether we compared Mexico to the US sample with the same nominal pay, b = 13.93, $t(2143) = 3.42$, $P < 0.001$, 95% CI 5.94 to 21.92, or that with same subjective pay, b = 19.28, $t(2167) = 4.71$, $P < 0.001$, 95% CI 11.26 to 27.30 (Fig. 3a and Appendix Table S3). Said differently, the effectiveness of monetary and non-monetary motivators was closer to one another in Mexico than in the US. In the US, money increased effort by 142.9% over the two pooled non-monetary conditions in the sample with the same nominal pay and by 165.6% in the sample with the same subjective pay (Fig. 3b). In Mexico, the difference was 41.6%.

Fig. 3: **Study 2b, Monetary Versus Pooled Non-Monetary Conditions in the US and Mexico.**



Effects of a monetary incentive (green) and pooled non-monetary treatments (flat fee and social norm; blue) in Mexico (N = 1,053 participants recruited on Prolific) and two samples in the US: one with the same nominal pay (N = 1,098 on Prolific) as in Mexico and one with the same subjective (Becker et al., 1964) pay (N = 1,122 participants recruited on Prolific) as in Mexico. Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive type in a multiple linear regression model is statistically significant, $b = 13.93$, $t(2143) = 3.42$, $P < 0.001$, 95% CI 5.94 to 21.92 for the comparison between Mexico and the US sample with the same nominal pay, and $b = 19.28$, $t(2167) = 4.71$, $P < 0.001$, 95% CI 11.26 to 27.30 for the comparison between Mexico and the US sample with the same subjective pay. Panel B shows the money advantage—that is, how much more effective money is than the pooled non-monetary treatments in each sample. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the monetary versus pooled non-monetary conditions.

Next, we analyzed individual conditions. The interaction between culture and incentive was statistically significant. The money advantage was larger in the US than in Mexico for the flat fee condition as comparison, b = -10.32, t(2141) = -2.24, P = 0.025, 95% CI -19.37 to -1.27, and for the norm condition as comparison, b = -17.69, t(2141) = -3.81, P < 0.001, 95% CI -26.80 to -8.58 (Additional Fig. 3a and Appendix Table S8). We found the same pattern in our comparison of Mexico and the US sample with the same subjective pay: money was relatively more effective in the US than in Mexico when compared to the flat fee condition, b = -13.41, t(2165) = -2.88, P = 0.004, 95% CI -22.55 to -4.27, and to the norm condition, b = -25.55, t(2165) = -5.47, P < 0.001, 95% CI -34.72 to -16.38.

Finally, we analyzed the amount of effort per dollar spent for each of the individual incentive conditions. In Mexico, the monetary condition was significantly less cost-effective than the norm condition, two-sided Welch's t(527.85) = -3.90, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = -9.03, d = -0.30, 95 CI% -13.58 to -4.49, but more cost-effective than the flat fee condition, two-sided Welch's t(605.56) = 4.00, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 8.07, d = 0.30, 95 CI% 4.11 to 12.03 (Additional Fig. 3c).

In both samples in the US, money was more cost-effective than the norm condition (Same Nominal Pay: two-sided Welch's t(671.05) = 5.09, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 10.97, d = 0.38, 95 CI% 6.74 to 15.20; Increased Pay: two-sided Welch's t(721.26) = 7.76, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 13.13, d = 0.57, 95 CI% 9.81 to 16.46) and the flat fee condition (Same Nominal Pay: two-sided Welch's t(665.00) = 13.70, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 21.85, d = 1.01, 95 CI% 18.72 to 24.99; Increased Pay: two-sided Welch's t(742.17) = 12.96, P < 0.001, P_{Bonf} < 0.001, $Mean_{\text{difference}}$ = 18.97, d = 0.95, 95 CI% 16.10 to 21.85). Thus, consistently with the findings in Study 2a, we found that a monetary incentive is

less cost-effective than a non-monetary intervention (social norm) in a "non-WEIRD" culture (Mexico), but not in a "WEIRD" culture (the US).

Study 2C: Pre-Registered Replication in the US and South Africa

In Study 2c, we made three improvements over Study 2b. First, we pre-registered the hypothesis and analysis. Second, we extended the sample to another "non-WEIRD" culture. We recruited Prolific workers in the US (two samples, N = 662 each; 318 female, 20 non-binary, $Mean_{Age}$ = 36.91, SD_{Age} = 12.62 in the sample with the same nominal pay; 323 female, 17 non-binary, $Mean_{Age}$ = 36.27, SD_{Age} = 12.52 in the sample with the same subjective pay) and South Africa (N = 649; 316 female, 6 non-binary, $Mean_{Age}$ = 28.29, SD_{Age} = 7.45). South Africa is an interesting test case because, although it scores lower on individualism than the US, it is more individualistic than China or India (Suh et al., 1998). Individualism is often contrasted with collectivism and refers to the tendency to prioritize individual goals and achievement (Markus, 2016; Oyserman et al., 2002; White & Lehman, 2005). South Africa's English-speaking population tends to be particularly individualist (Adams et al., 2012; Eaton & Louw, 2000). All participants in South Africa completed the study in English and reported speaking it fluently.

Third, we tested a wider range of psychological interventions. We tried to "stress test" our findings by choosing non-monetary interventions we suspected might be particularly effective in "WEIRD" cultures. In one condition, we told participants they were competing with other participants, since some researchers see competition as a facet of individualism (Triandis et al., 1988; Triandis & Gelfand, 1998). In another condition, we incentivized people with donations to organized charities, which receive more contributions in "WEIRD" compared to "non-WEIRD" cultures even after controlling for income (Einolf, 2017; Wiepking, 2021). Given

these stress tests, we anticipated smaller effect sizes than in Studies 2a and 2b. We compared these two conditions (competition and charity) with the monetary condition.

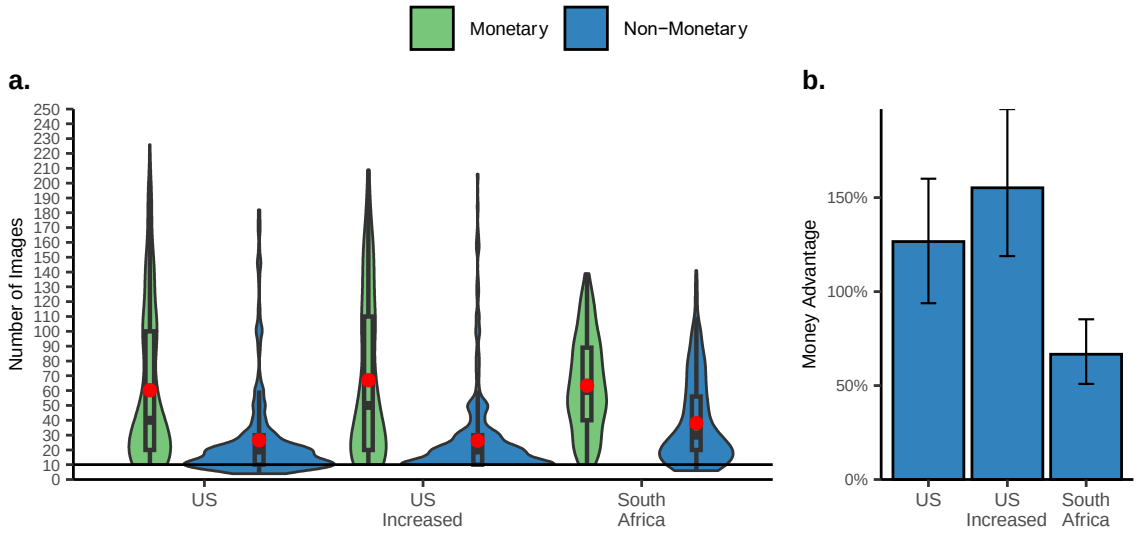
As in Study 2b, we ran some US participants with the same pay as offered to the participants in South Africa. Other US participants worked for pay adjusted to reflect local differences in purchasing power, using a procedure analogous to the one in Study 2b (Becker et al., 1964). In the US sample with the same subjective pay, we increased the base pay from $1.30 to $2.25 in all conditions as well as the piece-rate from 5 to 9 cents per 10 images in the monetary condition.

We pre-registered pooled analyses because we did not have sufficient power to compare individual conditions. We report the comparisons between the monetary and each of the non-monetary conditions in section "Mean Effort and Cost-Effectiveness in Individual Conditions in Study 2c" of the Appendix, but readers should be aware that the statistical power for these comparisons is lower.

Results

The difference in the effectiveness between monetary and non-monetary conditions was larger in the US than in South Africa, both when we compared the sample in South Africa to the US sample with the same nominal pay, $b = 7.98$, $t(1303) = 1.97$, $P = 0.049$, 95% CI 0.03 to 15.92, and to the US sample with the same subjective pay, $b = 15.40$, $t(1303) = 3.74$, $P < 0.001$, 95% CI 7.33 to 23.48 (Fig. 4a and Appendix Table S4). Additional money increased effort by 66.7% in South Africa, but by 126.6% in the US sample with the same nominal pay and by 155.2% in the US sample with increased pay (Fig. 4b).

Fig. 4: **Study 2c, Monetary Versus Pooled Non-Monetary Conditions in the US and South Africa.**



Effects of a monetary incentive (green) and pooled non-monetary treatments (competition and charity; blue) in South Africa (N = 649 participants on Prolific) and two samples in the US: one with the same nominal pay (N = 662 on Prolific) as in South Africa and one with the same subjective (Becker et al., 1964) pay (N = 662 participants on Prolific) as in South Africa. Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive type in a multiple linear regression model is statistically significant, $b = 7.98$, $t(1303) = 1.97$, $P = 0.049$, 95% CI 0.03 to 15.92 for the comparison between South Africa and the US sample with the same nominal pay, and $b = 15.40$, $t(1303) = 3.74$, $P < 0.001$, 95% CI 7.33 to 23.48 for the comparison between South Africa and the US sample with the same subjective pay. Panel B shows the money advantage—that is, how much more effective money is than the pooled non-monetary treatments in each sample. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the monetary versus pooled non-monetary conditions.

Studies 2a–c extended the findings from Study 1 to four other cultures and to a more meaningful task. The money advantage was higher in "WEIRD" than "non-WEIRD" cultures both when the contractual minimum was left ambiguous (Study 1) and when it was made explicit (Studies 2a–c). Cultural differences persisted after we adjusted the pay to reflect local purchasing power (Studies 2b and 2c).

Although we use the terms "WEIRD" and "non-WEIRD" throughout the paper, we agree with the researchers who coined this term that the underlying cultural differences are not binary (Graham et al., 2016; Henrich et al., 2010; Muthukrishna et al., 2020; Schulz et al., 2018). One way to move from the "WEIRD"–"non-WEIRD" dichotomy into a more nuanced continuum is the calculation of cultural distance. Researchers used responses to the World Values Survey to calculate different cultures' "distance" from the US (Muthukrishna et al., 2020). Interestingly, these cultural distance scores follow the strength of the money advantage in our studies (Additional Fig. 4). Future studies with more cultures and identical conditions can more robustly test the predictive power of cultural distance and other cultural and institutional dimensions, an idea we come back to in Study 5.

STUDIES 3A–B: COMPARING PSYCHOLOGICAL INCENTIVES AND MINIMAL PAY

In Study 3a, we pushed the boundaries of just how little extra money beyond the base pay it would take to spur extra effort. The bonuses we paid in Studies 2a–c were not large, but they were above average for crowdsourcing sites. Surveys have found that crowdsourcing workers average $3.31 per hour, and the pay is even lower for non-Western workers (Berg et al., 2018; Rani & Furrer, 2021). Our studies maxed out at 10 minutes, which would work out to 55 cents at

these sites' average wage. In the monetary conditions in Studies 2a–c, participants could earn more than that from the piece-rates alone (and the base pay made the earnings even better). Since the piece-rates were above average, it may not be surprising that monetary incentives would be more motivating than psychological interventions.

This made us wonder whether it was the actual pragmatic value of the money that motivated people or whether money holds a symbolic value beyond the actual usefulness of the amount. To test this, we changed the piece-rates in Study 3a to negligible amounts.

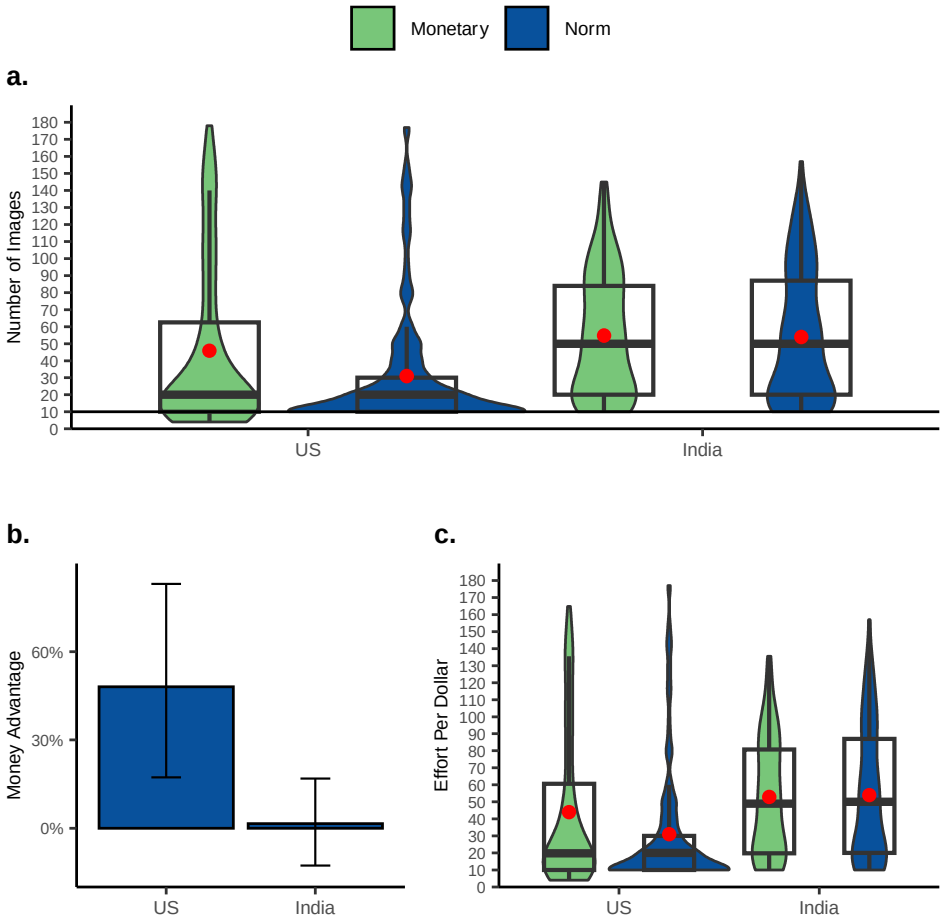
Study 3A: Lowering the Piece-Rate in India and the US

In Study 3a, we lowered the bonus to a penny for every 20 images completed. We compared this minimal pay condition to a social norm intervention in samples in India (N = 352 recruited on MTurk; 83 female, $Mean_{Age}$ = 35.79, SD_{Age} = 8.73) and the US (N = 382 recruited on Prolific; 197 female, 14 non-binary, $Mean_{Age}$ = 37.76, SD_{Age} = 13.99). We pre-registered the hypothesis, methods, and analysis. As in the previous studies, we analyzed the data using regressions controlling for age, gender, and education.

Results

The money advantage was larger in the US than in India, as evidenced by a statistically significant interaction between culture (US, India) and incentive (norm, monetary), b = 13.77, $t(726)$ = 2.31, P = 0.021, 95% CI 2.09 to 25.45 (Fig. 5a and Appendix Table S5). A tiny monetary incentive, when compared to the social norm condition, increased effort by just 1.6% in India but by 48.1% in the US (Fig. 5b).

***Fig. 5*: Study 3a, Norm Versus Minimal Pay Condition in India and the US.**



Effects of a minimal monetary incentive (of 1 cent per 20 image-ratings; green) and a social norm condition (blue) and in the US (N = 382 participants recruited on Prolific) and India (N = 352 participants recruited on MTurk). Panel A shows the central tendency and distribution of effort by incentive and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive in a multiple linear regression model is statistically significant, $b = 13.77$, $t(726) = 2.31$, $P = 0.021$, 95% CI 2.09 to 25.45. Panel B shows the money advantage—that is, how much more effective the minimal incentive is compared to the social norm condition in each country. Panel C shows the central tendency and distribution of cost-effectiveness (effort per dollar spent) of each incentive by country. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. Minimal monetary incentive is significantly more cost-effective than the social norm condition in the US, two-sided Welch's $t(365.20) = 3.14$, $P = 0.002$, $P_{Bonf} = 0.004$, $Mean_{\text{difference}} = 13.00$, $d = 0.32$, 95 CI% 4.86 to 21.15, but not in India, two-sided Welch's $t(345.72) = -0.27$, $P = 0.785$, $P_{Bonf} = 1.000$, $Mean_{\text{difference}} = -1.06$, $d = -0.03$, 95 CI% -8.73 to 6.60. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the minimal-monetary-incentive versus social-norm condition.

In India, paying 1 additional cent per every 20 images rated was not significantly more cost-effective than emphasizing a descriptive norm, two-sided Welch's $t(345.72) = -0.27$, $P = 0.785$, $P_{Bonf} = 1.000$, $Mean_{difference} = -1.06$, $d = -0.03$, 95 CI% -8.73 to 6.60 (Fig. 5c). In the US, minimal pay was significantly more cost-effective than the social norm condition, two-sided Welch's $t(365.20) = 3.14$, $P = 0.002$, $P_{Bonf} = 0.004$, $Mean_{difference} = 13.00$, $d = 0.32$, 95 CI% 4.86 to 21.15.

Therefore, Study 3a showed that relative to a psychological intervention, even a minimal monetary incentive was more motivating in a "WEIRD" compared to a "non-WEIRD" culture.

Study 3B: Is Minimal Pay More Effective Because of Gamification?

One alternative explanation for the findings in Study 3a is that participants in the US found a tiny monetary incentive more motivating not because it was money, but because of gamification—triggering the sense of competitive games or accumulating rewards gamification (Bouchrika et al., 2021; Koivisto & Hamari, 2019; Morschheuser & Hamari, 2019). Many smartphone apps take advantage of gamification to motivate people to rate restaurants, tag photos, or correct errors by rewarding points, even without money (Morschheuser & Hamari, 2019). We tested this alternative explanation in Study 3b.

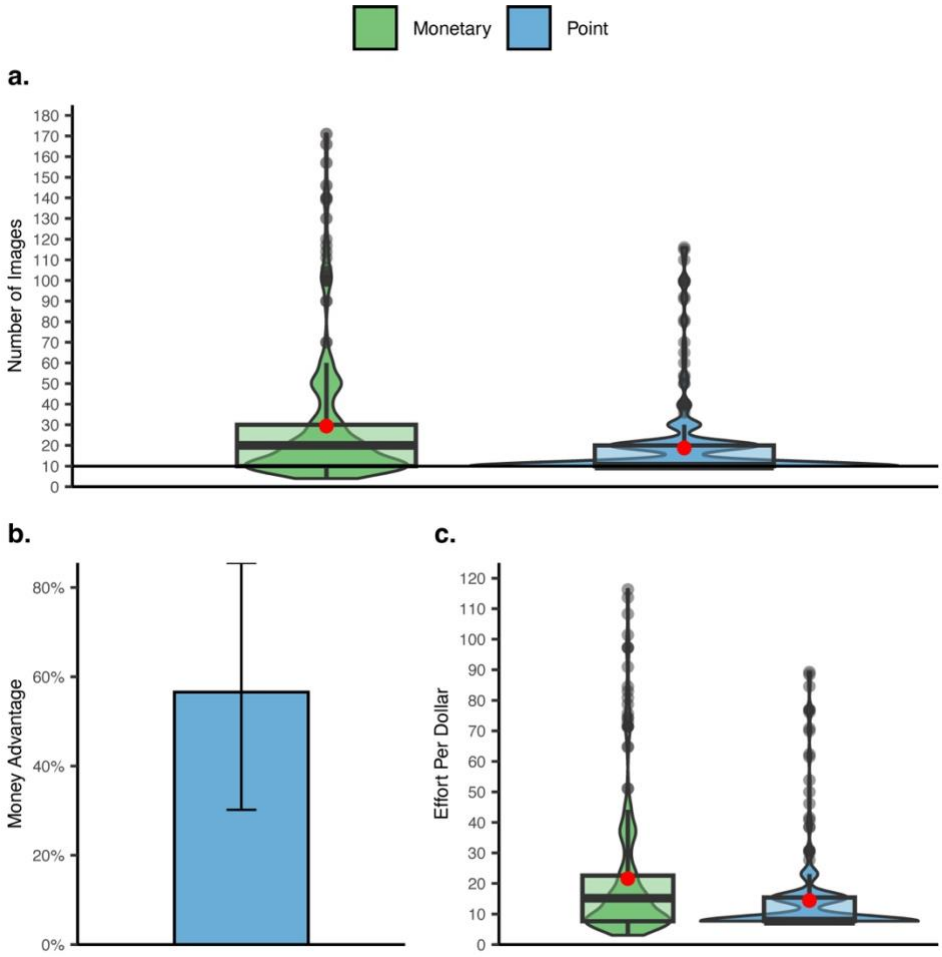
We pre-registered the hypothesis, methods, and analysis. We tested whether people in the US ($N = 537$ recruited on Prolific) would exert the same amount of effort when we rewarded them for rating each 10 additional pictures with a cent (minimal pay) or a point without any monetary value (gamification).

Results

Participants worked harder in the minimal pay ($M = 29.38$, $SD = 31.90$) than in the gamification condition ($M = 18.76$, $SD = 19.39$), Welch's $t(446.98) = 4.67$, $P < 0.001$,

$Mean_{difference}$ = 10.61, d = 0.40, 95% CI 6.15 to 15.08 (Fig. 6). They completed 56.6% more pictures when working for cents as compared to for points.

***Fig. 6*: Study 3b, Minimal Pay Versus Points (Gamification) Condition in the US.**



Effects of the minimal monetary incentive (1 cent per rating 10 images; in green) and a non-monetary gamification treatment (1 extra point per rating 10 images; in blue) in the US (N = 537 on Prolific). Panel A shows the central tendency and distribution of effort by incentive condition. The black line within each box represents a median; the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The difference in the number of images rated in the two conditions is statistically significant, Welch's $t(446.98)$ = 4.67, $P < 0.001$, $Mean_{difference}$ = 10.61, d = 0.40, 95% CI 6.15 to 15.08. Panel B shows the money advantage—that is, how much more effective the minimal monetary incentive is compared to the gamification condition. Panel C shows the central tendency and distribution of cost-effectiveness (effort per dollar spent) of each incentive. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The minimal monetary incentive is more cost-effective than the gamification treatment, Welch's $t(471.93) = 4.37$, $P < 0.001$, $Mean_{difference} = 7.15$, d = 0.38, 95% CI 3.93 to 10.36. In Panel B, the error bar is a bootstrapped 95% CI for the mean relative difference in the number of images rated in the minimal-monetary-incentive.

The minimal pay condition was more cost-effective than the gamification condition, Welch's $t(471.93) = 4.37$, $P < 0.001$, $Mean_{difference} = 7.15$, $d = 0.38$, 95% CI 3.93 to 10.36 (Fig. 6c).

The results of Study 3b show that gamification cannot not explain the effectiveness of tiny amounts of money in Study 3a. Instead, participants in the US seem to be motivated specifically by money.

STUDY 4: RANDOMLY ASSIGNING CULTURAL FRAMES THROUGH LANGUAGE

In Study 4, we conducted a lab-in-the-field experiment to overcome the problem of causality. In interpreting the results of our previous studies, we attributed the differences to culture, but we did not randomly assign culture. To get at this issue, we randomly assigned bilingual people in India (N = 2,065; 286 female, 2 non-binary, $Mean_{Age} = 24.87$, $SD_{Age} = 4.90$) to complete the study in Hindi or in English.

Previous studies have found that having bilingual people switch languages activates a discrete set of social and moral norms in the culture associated with each language (Benet-Martínez et al., 2002; Luna et al., 2008; Ralston et al., 1995; Trafimow et al., 1997; Verkuyten & Pouliasi, 2002). For example, one study surveyed managers in Hong Kong in Chinese or English (Ralston et al., 1995). When surveyed in English, the surveyed managers endorsed values more similar to those espoused by American managers. For instance, they rated conformity and tradition as less important, while achievement and hedonism as more important.

We recruited participants through a post on the Facebook group QMaths. This group has over 280,000 members interested in preparing for competitive exams for jobs in sectors ranging from banking to railways.

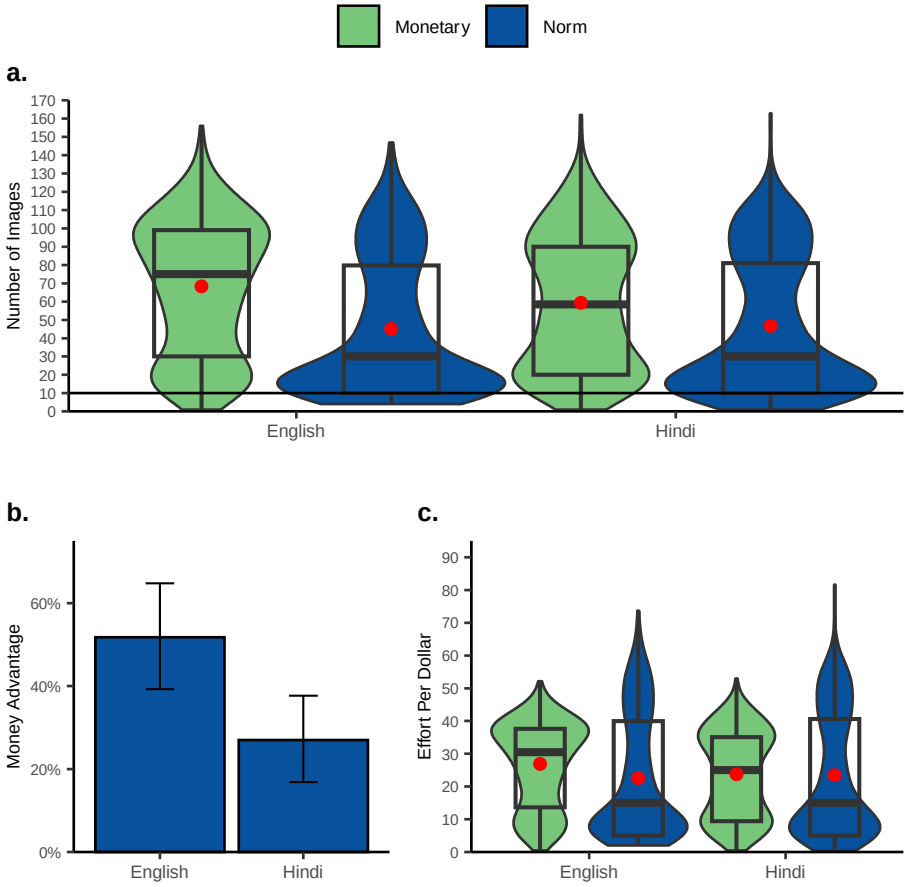
We compared how many images participants completed when given a monetary reward versus a psychological intervention (social norm). In addition, we asked questions to explore the mechanism behind the cultural differences. The seven exploratory variables asked about participants' motivation for and perception of the task, such as whether they enjoyed completing the task and whether they completed it only for money.

Results

The money advantage was larger when participants took the study in English than in Hindi, as evidenced by an interaction between language and incentive type, $b = 10.69$, $t(2061) = 3.31$, $P = 0.001$, 95% CI 4.35 to 17.02 (Fig. 6a and Appendix Table S6). The difference in effort between the monetary and the norm condition was 51.8% in English and 27.0% in Hindi (Fig. 6b).

Post-hoc tests (with Bonferroni corrections) showed that money was significantly more motivating in English than in Hindi, two-sided Welch's $t(1030.41) = 3.94$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 8.94$, 95% CI 4.48 to 13.40, 97.5% CI 3.84 to 14.04, $d = 0.24$. On the other hand, people assigned to the non-monetary condition that highlighted the descriptive norm rated the same number of images in English and in Hindi, two-sided Welch's $t(1028.98) = -0.760$, $P = 0.447$, $P_{Bonf} = 0.894$, $Mean_{\text{difference}} = -1.74$, $d = -0.05$, 95% CI -6.25 to 2.76, 97.5% CI -6.90 to 3.41.

Fig. 7: **Study 4, Norm Versus Monetary Condition in India, by Language Prime (English or Hindi).**



Effects of a monetary incentive (green) and a social norm treatment (blue) in India (N = 2,065 participants recruited on Facebook), by assigned language (English or Hindi). Panel A shows the central tendency and distribution of effort by language and incentive conditions. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between language and incentive in a multiple linear regression model is statistically significant, $b = 10.69$, $t(2061) = 3.31$, $P = 0.001$, 95% CI 4.35 to 17.02. Panel B shows the money advantage—that is, how much more effective the monetary condition is compared to the social norm condition. Panel C shows the central tendency and distribution of the cost-effectiveness (effort per dollar spent) by language and incentive. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The monetary incentive is more cost-effective than the social norm condition in English, two-sided Welch's $t(919.21) = 4.37$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{difference} = 4.34$, $d = 0.27$, 95% CI 2.39 to 6.28, but the two incentives do not statistically significantly differ in their cost-effectiveness in Hindi, two-sided Welch's $t(915.62) = 0.30$, $P = 0.761$, $P_{Bonf} = 1.000$, $Mean_{difference} = 0.30$, $d = 0.02$, 95% CI -1.64 to 2.24. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the norm versus monetary condition.

In our analyses on cost-effectiveness, we found that the monetary incentive was more cost-effective than the social norm in English, two-sided Welch's $t(919.21) = 4.37$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{difference} = 4.34$, $d = 0.27$, 95% CI 2.39 to 6.28 (Fig. 6c). By contrast, in Hindi, the two conditions did not significantly differ in their cost-effectiveness: two-sided Welch's $t(915.62) = 0.30$, $P = 0.761$, $P_{Bonf} = 1.000$, $Mean_{difference} = 0.30$, $d = 0.02$, 95% CI -1.64 to 2.24.

Next, we analyzed whether the cultural prime influenced motivations. Language produced a statistically significant difference on one of seven items (with Bonferroni corrections). Participants were more likely to say that they completed the task only for money in English than in Hindi, $b = 0.40$, $t(2017) = 3.49$, $P = 0.001$, $P_{Bonf} = 0.004$, 95% CI 0.18 to 0.63, 99.3% CI 0.09 to 0.71, which suggests that English activated a more transactional mental frame across incentive treatments. We report further details on the motivation questions in the Appendix Tables S32, S33a, S33b, S34 and Fig. S4.

Thus, in Study 4, we found that English increased the money advantage. However, we recognize that using language to switch cultural frames is not reducible to the simple claim that the Hindi condition "randomly assigns" Indian culture, while the English condition "randomly assigns" US (or UK, or "WEIRD") culture. The differential set of associations activated by the two languages is likely to be considerably more complex (Benet-Martínez et al., 2002; Hong et al., 2000). For instance, English might activate concepts related to work, professional life, or achievement. This may be particularly true given that English is the primary language used for higher education in India and is often associated with status and prestige (Altbach, 2009; Si, 2011). Similarly, people who completed the study in English may have implicitly compared themselves to native English speakers, which could have led to more upward social comparisons. These are empirically testable conjectures worthy of unpacking in future studies. However, it is

worth remembering that professionalism, drive for prestige, and social comparison are all part of culture too.

STUDY 5: GOING BEYOND THE CONTRACT IN 40 COUNTRIES

Does the cultural variation in the money advantage show that money is more effective in "WEIRD" cultures, psychological interventions are more effective in "non-WEIRD" cultures, or both? In Studies 2a–c, the answer seems to depend on each individual pair of countries. For example, in South Africa and the US, participants rated the same number of images in the monetary condition, and the effect seemed to come from non-monetary incentives being more effective in South Africa. On the other hand, participants in China did not seem to be more motivated than people in the UK by psychological interventions. Instead, money being more motivating in the UK than in China seemed to produce the interaction. Similarly, in Study 4, money seemed to be driving the effect of the language frames on effort.

We argue that when populations differ in the level of baseline motivation, making absolute claims about the motivating force of money may be misleading, and we do not believe that our results show that money is inherently less motivating to "non-WEIRD" workers. Consider the finding by Gneezy and colleagues who showed that monetary incentives improved students' performance on a math test in the US but not in China (Gneezy et al., 2019). Can we conclude that money is less motivating for young people in China?

In both countries, the study had to compare the effectiveness of added money to the set of motivators already present, such as recognition or disapproval from teachers and parents. Students in China might have been more motivated to perform well to begin with, and adding a

financial incentive could only raise their effort a little, if at all. This makes it difficult to conclude that money—and nothing else—is what produced the cultural difference.

In our studies, the "control" condition was also not a true blank slate: we should not assume that people across cultures and levels of economic security behave in the same way when working for a flat fee. Instead, the likelihood of doing more than what is minimally required and how one interprets that minimum represent meaningful cultural differences in psychological contracts and in motivation more broadly (Bass, 1997; Falk et al., 1999).

In Studies 5a and 5b, we measured effort in the absence of explicit monetary incentives in 40 countries to examine what motivates workers to go above and beyond the minimum amount of effort required from them to receive pay.

Study 5A: Doing More When the Absence of Monetary Reward Is Explicit

In Study 5a, we recruited 7,823 participants in 40 countries across seven crowd-sourcing platforms to work on a data entry task. This study served as a way to get acquainted with several crowd-sourcing platforms that are less commonly used by social and behavioral scientists and as a pilot study for a project that is not part of this dissertation.

Participants had to transcribe a hand-written sheet that contained details of previous orders of an anonymous business (Appendix Image S3). Like in previous studies, workers received a flat-fee wage, which ranged between $0.25 and $2.50 depending on country and crowd-sourcing platform, with workers in "WEIRD" countries being paid, on average, 5 times more than workers in "non-WEIRD" countries.

Again, we established an explicit contractual minimum; this time, it concerned the minimal amount of time that workers had to spend on the task (5 minutes) as opposed to the amount of output they had to produce like in our previous studies (10 images). Specifically, we

told workers, “You will not get paid more if you spend more than 5 minutes. Regardless of how much time you spend…you will be paid exactly the same amount of money.”

By design, the sheet was nearly impossible to complete within five minutes. Our primary measure of effort was the proportion of people per country who completed the entire data entry task, going beyond the minimal amount of effort required. However, as secondary measures of effort, we also analyzed the total number of cells inputted and total number of cells inputted correctly (both out of 132). The task completion measures correlated $r = 0.89$ with the two continuous effort measures, which themselves correlated $r = 0.96$ with each other, suggesting that most cells inputted by the workers in our studies were inputted correctly.

Attention and Comprehension Checks

We implemented strict checks to ensure participants understood the task. First, we asked workers about their understanding of the explicit contractual minimum: they had to correctly state how much effort they had to exert to receive the flat-fee wage. Those who failed to answer correctly did not proceed to the data entry task. Full text of the instructions and the comprehension check is presented in the Methods section.

Second, we asked workers, after the completed the data entry task, “What did the instructions say about whether you would not or would not receive more pay for spending more than 5 minutes on the data entry task?” Those who said that the instructions had stated that workers “would receive more pay for spending longer than 5 minutes” were excluded from the sample. Because of this, we are able to assert that potential differences in expected personal benefits in our samples (e.g., the possibility that the employer reciprocates higher effort with a bonus) do not stem from people in “non-WEIRD” countries being more likely to misunderstand the instructions.

Finally, we excluded workers who, when asked to select a specific value on a scale toward the end of the study, selected a value different from the one we requested. The total sample included data from 7,823 participants recruited on seven crowd-sourcing platforms in Study 5A. As before, we summarize the exclusions in the section "Exclusion Data and Criteria for Studies 2–5" of the Appendix. We report the proportions of participants in each country who failed each of the checks in Appendix Table S29f: "Sample Characteristics for Studies 5A and 5B."

Completion Speed Pre-Test

In a pre-registered pre-test, reported in the Appendix Section "Task Completion Speed Pre-Test for Study 5A" and in the Appendix Fig. S3, we incentivized workers in five countries—Egypt, Turkey, Poland, the UK, and the US—with monetary incentives to transcribe the entire data entry sheet as quickly and accurately as they could. We found that 2.9% (17 out 578) of people were able to complete the task in 5 minutes or less.

We also confirmed that workers in Western countries tend to be faster at data entry. UK workers completed the task in around 11 minutes, while Turkish workers took over 3 minutes longer on average, and Egyptian workers need nearly twice more time, with an average completion time of around 21 minutes. Because we also incentivized accuracy (that is, ensuring that every cell in the data entry sheet is inputted correctly), these times might underestimate the times in Study 5a, where the instructions did not mention accuracy.

Controlling for Device Used

Importantly, one reason why workers in Western countries had faster average completion times in the pre-test was that more workers in Western countries worked on laptops. World Bank estimates that fewer than 20 percent of households in Nigeria or Myanmar owned a computer,

compared to over 80 percent in Australia or Japan (World Bank, 2024). These figures align with the rates of using laptops in our sample, displayed in Fig. S5 and Table S29e of the Appendix.

In Study 5a, we could not restrict all workers to using laptops, because most crowd-sourcing sites, unlike platforms like Prolific or MTurk, do not have this feature. To account for differences in device use, we took the following three steps. First, we controlled for device type both as a fixed effect and as a random effect in the regressions, unless the random effect for laptop use was perfectly correctly with that of country and was thus redundant. Second, to have enough data to more accurately adjust for device use, we recruited more workers in countries—predominantly "WEIRD"—where few participants used a phone—if the platform where we recruited participants in those countries allowed for such selective recruitment.

Third, for illustrative purposes, we display marginal proportions that adjust for device type and present the results as if the samples had the same composition of phone and laptop users. This simple transformation allows for more accurate comparisons between countries like Nigeria, where only 1 in 4 people in our sample worked on a laptop, and countries like Japan, where 9 in 10 people did.

Figure S5 in the Appendix shows how marginal completion rates differ from the raw proportions based on the rate of laptop use in a given country. For instance, compare Egypt, where 37.2% of the participants worked on a laptop, to Morocco, where 63.5% of workers did. The "raw" completion rates for both are nearly identical, with Morocco slightly ahead (46.2% in Egypt vs. 46.4% in Morocco). However, among workers who used laptops, workers in Egypt were more likely to complete the task than workers in Morocco (64.4% in Egypt versus 53.2% in Morocco), and the two countries did not differ in how likely those who worked on their phones were to complete the task (34.5% in Egypt versus 34.4 % in Morocco). Thus, the reason why

Morocco had a 0.2% higher rate of task completion when we consider raw proportions is due to the mere fact that more people in Morocco worked on laptops. The marginal proportions adjust for this difference in device use between the two countries and place Egypt five percentage points ahead of Morocco. Additional Fig. 6 and Appendix Tables S12 and S13 display raw proportions separately for each device type (laptop or phone).

Measure of Cultural Difference

As a measure of difference in socio-cultural environments, we used the three-level "WEIRD" country classification, which was adapted from Talhelm and colleagues (under review). It categorized countries as "non-WEIRD," "Partly (Sometimes) WEIRD," or "WEIRD," based on their economic development, cultural history, and prevalent psychological characteristics such as individualism, and was coded as an ordinal variable with three levels. The only change to the classification made in this dissertation was to categorize Japan as "Partly (Sometimes) WEIRD" and not as "non-WEIRD," due to its high level of economic development, G7 (the Group of Seven) membership and data that shows Japan is more individualistic than other East Asian countries (Minkov et al., 2017; Takano & Sogon, 2008).

Analysis

We used a mixed-effects logistic regression models to predict the binary outcome of task completion (completed or not). The models included gender, age, employment status, student status, education level, urban living, majority race, socioeconomic status, income, platform, and device type, while allowing the effect of device type to vary across countries.

Compared to the simpler model that only included different intercepts for countries, including a random slope helped us account for the possibility that the influence of using a laptop on task completion may not be the same across countries: in some countries, using a laptop

might significantly boost task completion rates compared to working on a phone, while in others, it might have a smaller effect or no effect at all.

For secondary dependent variables, we used the same predictors but employed a mixed-effects linear regression model to handle continuous outcomes like the total number of cells inputted. For all regressions in Study 5, all the predictors are scaled, allowing one to interpret standardized coefficients as effect sizes.

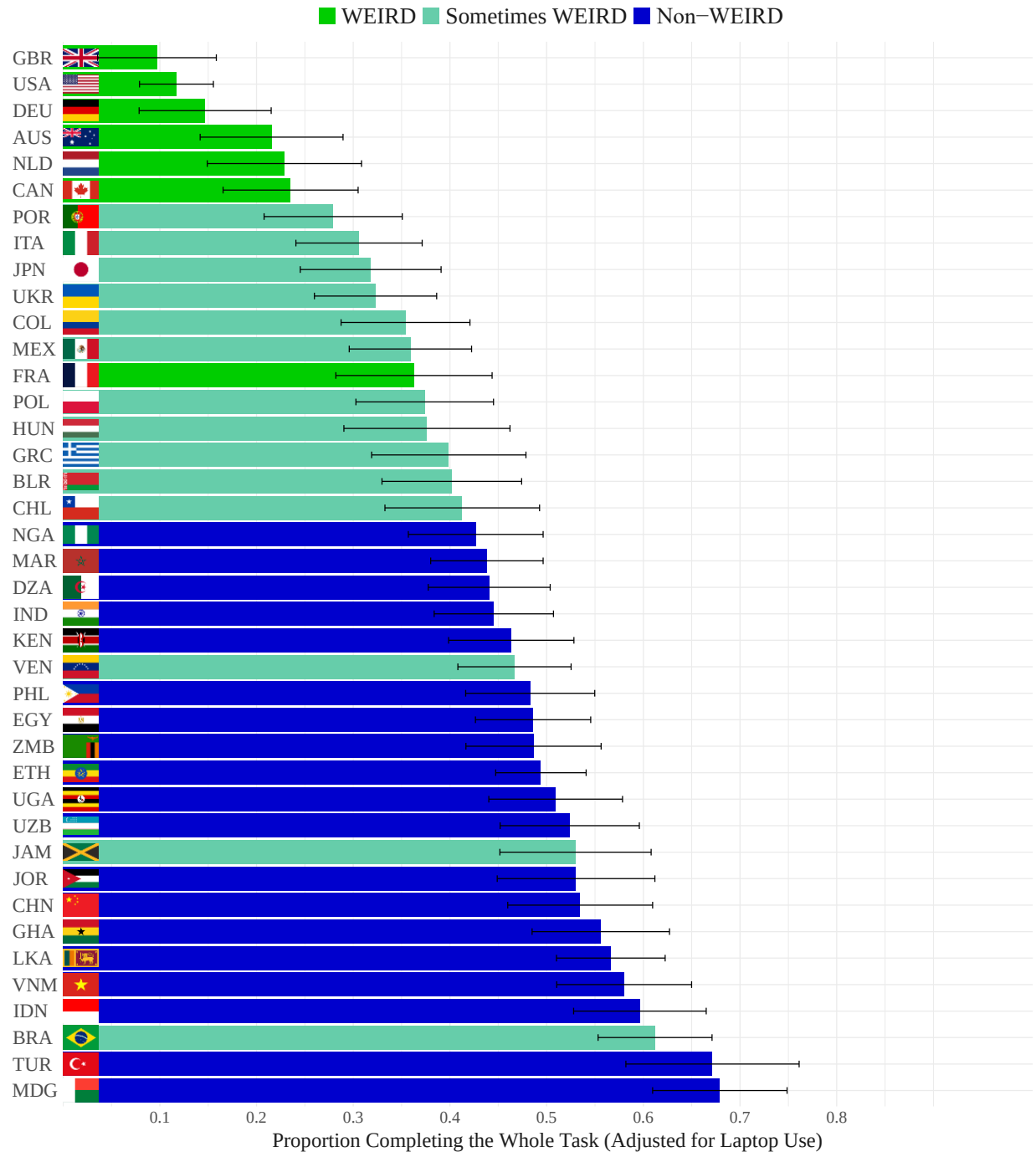
Results

Proportion Completing the Whole Task

Twenty-two percent of workers in "WEIRD" countries completed the whole task, compared to 44% in "Partly (Sometimes) WEIRD" and 52% in "non-WEIRD" countries (Fig. S7b in the Appendix). Completion rates for individual countries ranged between around 10% in the UK to around 70% in Turkey and Madagascar (Fig. 8).

The regression analysis showed a significant negative linear effect of the "WEIRD" country classification on the probability of completing the whole task, b = -1.035, z = -6.070, P < 0.001, 95% CI -1.370 to -0.701, indicating that participants from more "WEIRD" countries were less likely to complete the entire transcription (Fig. 8). The quadratic term for the "WEIRD" classification was also significant and negative effect, b = -0.315, z = -2.494, P = 0.013, 95% CI -0.562 to -0.067, suggesting the decline in completion rates among "WEIRD" countries compared to "non-WEIRD" and "Partly (Sometimes) WEIRD" countries was not entirely linear.

Fig. 8: **Study 5a, Proportion of Participants per Country Who Complete the Whole Task.**



Proportion of participants in Study 5a (*N* = 7,823 independent participants recruited on seven crowd-sourcing platforms) who completed the entire data entry task when doing so was not contractually required and would not have resulted in a monetary reward, as per explicit instructions. The figure presents the marginal means for completion rates adjusted for laptop use, colored by "WEIRD" country classification adapted from Talhelm and colleagues (under review). The mixed-effects logistic regression analysis showed a significant negative linear effect of the "WEIRD" country classification on the probability of completing the whole task, b = -1.035, z = -6.070, P < 0.001, 95% CI -1.370 to -0.701, indicating that participants from more "WEIRD" countries were less likely to

complete the entire transcription. The quadratic term for the “WEIRD” classification was also significant and negative, b = -0.315, z = -2.494, P = 0.013, 95% CI -0.562 to -0.067, suggesting a non-linear decline in completion rates along the “WEIRD” classification. Error bars are 95% CIs.

Continuous Measures of Effort

Analyzing the total cells inputted in a mixed-effects linear regression, we also found a significant negative linear effect of the "WEIRD" country classification, b = -0.374, z = -4.904, $P < 0.001$, 95% CI -0.524 to -0.225, showing that people from more "WEIRD" countries completed fewer cells (Additional Fig. 7). The quadratic term for the "WEIRD" classification also had a significant negative effect, b = -0.132, z = -2.520, P = 0.016, 95% CI -0.234 to -0.029, suggesting a more pronounced decline in total number of cells inputted in "WEIRD" countries compared to "non-WEIRD" and "Partly (Sometimes) WEIRD" countries.

For cells inputted correctly, both the linear and the quadratic effects were similarly negative and significant, b = -0.385, z = -5.116, $P < 0.001$, 95% CI -0.533 to -0.238 for the linear term, and b = -0.135, z = -2.628, P = 0.012, 95% CI -0.236 to -0.034 for the quadratic term.

Thus, across all three measures of effort, people from more "WEIRD" countries were more likely to go above the minimum required of them to get paid, and this decline was not linear among the levels of the "WEIRD" classification.

Moral Evaluations of Working Beyond the Contractual Minimum

Next, we analyzed the factors that correlated with the differences in effort between workers from more and less "WEIRD" countries. First, workers in non-Western countries were more likely to see working beyond the contractual minimum as the morally right thing to do. The linear term for "WEIRD" showed a significant negative effect, b = -0.211, t = -3.611, P = 0.001, 95% CI -0.326 to -0.097, indicating that individuals from "WEIRD" countries were less likely to agree that it is morally right to complete the task (Additional Fig. 8 and Appendix Table S20). The quadratic term was not significant, b = -0.001, t = -0.028, P = 0.978, 95% CI -0.079 to 0.076, suggesting that a decrease in moral evaluations was not steeper for "WEIRD" or for

"Partly (Sometimes) WEIRD" countries above the linear effect. These differences held when controlling for completion rates across countries, as participants who completed the whole task (vs. not) were significantly more likely to view it as the morally right thing to do, $b = 0.707$, $t = 27.203$, $P < 0.001$, 95% CI 0.657 to 0.758.

Are "non-WEIRD" Workers More Hard-Working and Prosocial?

Higher effort on the part of our "non-WEIRD" workers does not mean that people in "non-WEIRD" countries exert effort when there is no explicit financial reward in a manner necessarily devoid of considerations of personal benefits (Kreps, 1997). For instance, the Chinese concept of *guanxi* and the Russian concept of *blat* refer to fostering a network of social connections—an "economy of favors," which can allow one to get access to scarce items or circumvent legal barriers (Ledeneva, 1998, 2008; Michailova & Worm, 2003). Performing a favor (*renqing*) builds trust (*xinren*) and results in an implicit expectation of reciprocation (Barnes et al., 2011; Jia & Zsidisin, 2014). We argue that people might similarly choose to work over the contractual minimum "as an investment toward the possibility of attaining desired outcomes from them in the future" (Baranowska-Rataj et al., 2023; Leary et al., 2014).

For "non-WEIRD" workers, especially those in lower-income countries, exerting effort beyond the minimum needed for pay might be more important, as hiring decisions in these contexts are often based less on meritocracy (Bandiera et al., 2024). Instead, employers in "non-WEIRD" countries often rely on network-based hiring (Akcigit et al., 2021; Chandrasekhar et al., 2020). For instance, they are often unwilling to recruit outside managers when their businesses grow, preferring to hire within family instead due to a lack of trust toward strangers (Akcigit et al., 2021; Bertrand & Schoar, 2006; Caria & Falco, 2022; Fukuyama, 1995).

In Study 5a, we measured how people evaluated those who did not complete the task. These measures, we argue, tap into the social incentive structures that determine the outcomes

for behaviors in each social environment: workers should be more motivated to go beyond the contractual minimum if their social environments condemn those who do not.

In the main text, we report the analyses of the social evaluations of those who only did the amount of work that was contractually required. Appendix Figures S9 and S10 and Tables 22 and 23 further present the data for additional measures of personal benefits, such as expectations of higher pay or receiving future work offers for doing more work than required.

Social Evaluations of Those Who Do Not Complete the Task

The negative social evaluation index comprised five items: whether people saw those who stopped working after 5 minutes and did not complete the entire transcription as 1) intelligent (reverse scored), 2) selfish, 3) lazy, 4) not nice, and 5) said they would not trust them.

The mixed-effects linear regression on the index showed a significantly negative linear term for the "WEIRD" classifications, indicating that participants from "WEIRD" countries were less likely to have negative social evaluations of those who did not complete the task, b = -0.160, t = -2.258, P = 0.027, 95% CI -0.298 to -0.021 (Additional Fig. 9 and Appendix Table S21). Conversely, the quadratic term for "WEIRD" was not significant, b = 0.008, t = 0.164, P = 0.870, 95% CI -0.091 to 0.108, suggesting that a decrease in social evaluations was not steeper for "WEIRD" or for "Partly (Sometimes) WEIRD" countries above the linear effect. Additional Fig. 9 shows the effects on each individual measure that makes up the index. For example, it demonstrates that workers in Germany and the US associate not completing the task—when the task does not offer additional pay—with intelligence.

The cultural differences in social evaluations held when we controlled for completion rates. As with the moral judgment question, participants who completed the whole task were significantly more likely to negatively evaluate those who did not complete the task, b = 0.828, t

= 31.871, $P < 0.001$, 95% CI 0.777 to 0.879. The interaction between task completion and the linear term for the "WEIRD" classification was significant, indicating that among people who completed the whole task, the negative social evaluations were less pronounced for those from more "WEIRD" countries, $b = -0.169$, $t = -3.700$, $P < 0.001$, 95% CI -0.259 to -0.080. Similarly, the interaction between task completion and the quadratic term for "WEIRD" was significant, suggesting that the slope was not identical across "WEIRD" and "Partly (Sometimes) WEIRD" cultures, $b = -0.152$, $t = -3.557$, $P < 0.001$, 95% CI -0.235 to -0.068.

(Mis)predictions of Effort

We also asked participants to estimate the completion rates in their countries, which we thought might tap into perceived descriptive norms for effort. We present the results in the Appendix Figures S11 and S12. The estimates of completion rates vary between 24% in the UK to 54% in China, which means that the predicted range across countries is lower than the range we see for actual mean completion rates unadjusted for device type (12% in the UK to 71% in Madagascar; Fig. S11).

We find that the majority of the workers underestimate completion rates. Only participants in the US, the UK, and Germany significantly overestimate the proportion of participants who completed the whole task in their respective countries. This underestimation of effort is largely driven by people who do not complete the entire task themselves considerably underestimating mean effort, thinking that their peers are similarly unlikely to do so (Fig. S11). Among people who do not complete the entire task, we see a significant underestimation in 30 out of 40 countries (37 out of 40 countries underpredict effort directionally). For example, people who do not complete the task in Turkey estimate that 29% of their compatriots finish the transcription, whereas the actual completion rate we observe in Turkey is 67%.

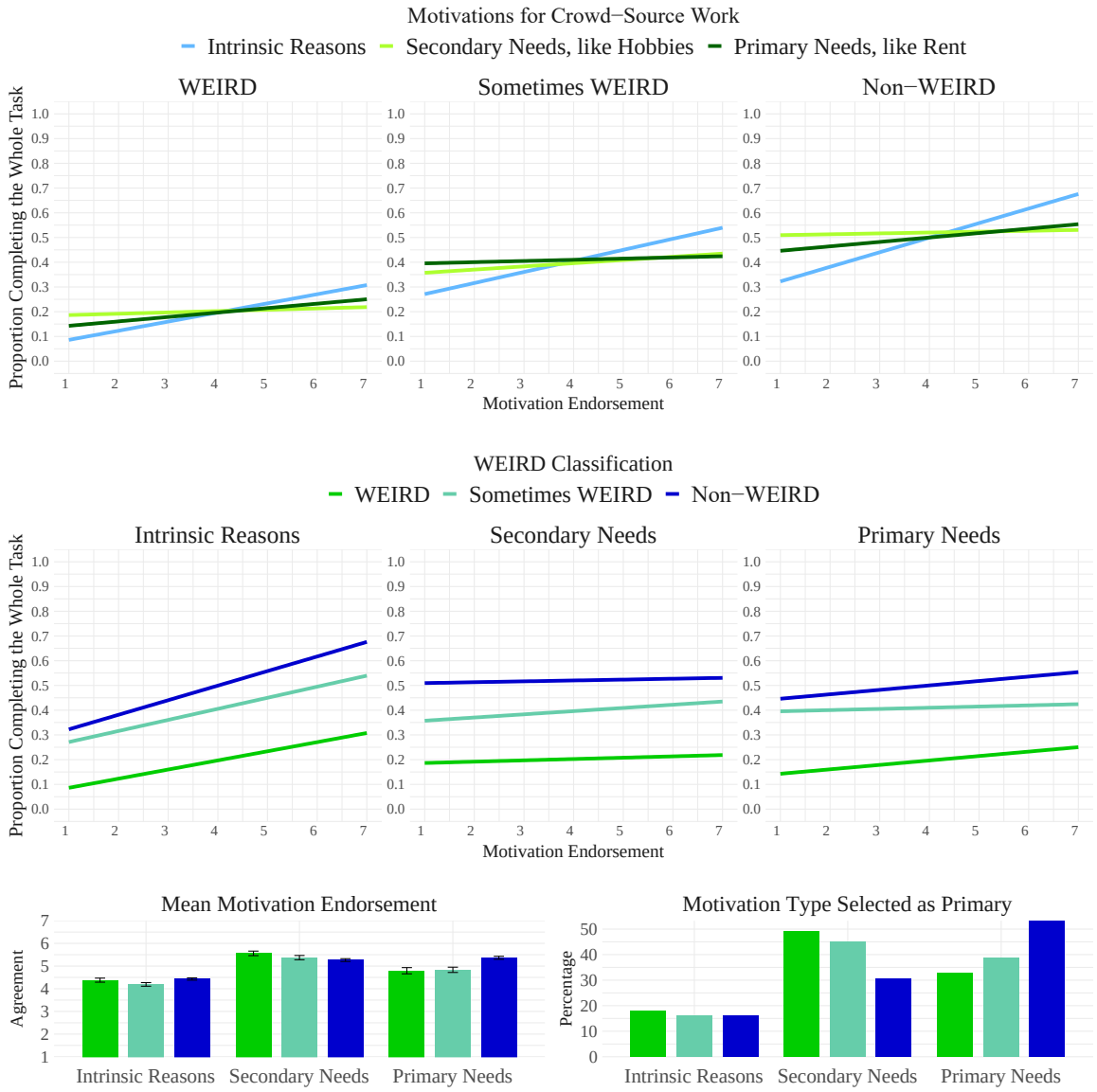
On the other hand, people who complete the entire transcription overestimate mean effort, with largest overestimations occurring in "WEIRD" countries. The minority of people who complete the whole data entry task in the US, UK, and Germany overpredict mean completion rates in their countries by over 30% (53% predicted versus 17% actual completion rate in the US, 48% versus 13% in the UK, and 54% versus 20% in Germany).

The Effect of Self-Reported Motivation for Doing Crowd-Source Work

Some participants (N = 4,368) in our sample were asked additional measures about their experience and motivations around crowd-sourcing. These measures included whether income from crowd-sourcing was their main source of income and whether they were satisfied with their wages (Fig. S13 in the Appendix).

In the main text, we analyze how completion rates differ based on participants' motives for work: intrinsic reasons, such as to kill time or benefit the society; secondary financial needs, such as earning money for hobbies or going out; and primary economic reasons, such as paying for food and rent. We asked these questions on 7-point Likert scales to get continuous measures of endorsement of different motives, which we then used in the reported regressions. We also asked a question that prompted people to select their single main reason for working on crowd-sourcing platforms. Fig. 9 presents both of these measures, grouped by "WEIRD" classification, and Appendix Fig. S14 reports country-level means for the binary choice measure.

Fig. 9: **Effect of Self-Reported Motivation Type on the Likelihood of Completing the Entire Task.**



Effects of three motivation types (Intrinsic, Secondary Monetary Needs, and Primary Monetary Needs) on the likelihood of completing task among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" countries among the proportion of participants in Study 5a who got asked additional measures about their experience and motivations around crowd-sourcing (N = 4,368 independent participants recruited on seven crowd-sourcing platforms). The top two figures present correlations between agreement with each motivation type (measured on a 7-point Likert scale) and the (raw) proportion completing the task. The bottom left figure shows mean agreement with each motivation type per "WEIRD" classification, where error bars represent 95% Cis. The bottom right figure shows percentage of participants per each "WEIRD" classification who select one of the three motivation types as their primary when answering the question, "Why do you complete tasks on Platform ("Platform" was replaced with each specific platform name)? Please select one main reason." The choices answer for the motivations type categories were "To earn money for food, rent, or bills," which we renamed "Primary Needs;" "To earn money for hobbies, gadgets, or going out," which we renamed "Secondary Needs;" and five answer choices that we merged into a category of "Intrinsic" reasons: "To kill time;" "Because the tasks are fun;" "To benefit society," "To learn new things," and "To share my opinions with someone who cares about what I think." The "WEIRD" classification

was adapted from Talhelm et al (under review). Being more motivated by primary economic needs showed a significant positive effect on task completion, $b = 0.127$, $z = 2.722$, $P = 0.006$, 95% CI 0.035 to 0.218, indicating a higher likelihood of completing the task. The effect of secondary needs was not significant, $b = -0.064$, $z = -1.326$, $P = 0.185$, 95% CI -0.158 to 0.030, suggesting no substantial influence on task completion. Intrinsic motivation demonstrated the largest significant positive effect, $b = 0.315$, $z = 6.499$, $P < 0.001$, 95% CI 0.220 to 0.410. The only significant interaction between "WEIRD" classification and motivation type was observed for the quadratic term of "WEIRD" classification and Primary Needs, suggesting that workers in more "WEIRD" countries who are primarily driven by earning money on necessities are more likely to complete the entire task, though the relationship is not linear along the levels of the "WEIRD" classification, $b = 0.166$, $z = 2.199$, $P = 0.028$, 95% CI 0.018 to 0.315.

When we analyzed how motive endorsement affects the likelihood of working above the contractual minimum, we observed two main patterns of findings. First, people driven by intrinsic reasons are more likely to complete the task, regardless of country, $b = 0.315$, $z = 6.499$, $P < 0.001$, 95% CI 0.220 to 0.410 (Fig 9), which aligns with previous research that found that people who are more motivated by factors beyond money, such as enjoyment, persist on tasks longer (Woolley & Fishbach, 2021).

Being motivated by primary economic needs similarly has a significant positive effect on task completion, though this effect is smaller, $b = 0.127$, $z = 2.722$, $P = 0.006$, 95% CI 0.035 to 0.218, indicating a higher likelihood of completing the task of completing the task, regardless of country, for those who primarily work on crowd-sourcing platforms to earn money on economic necessities. Conversely, being motivated by secondary economic needs was not significantly associated with task completion, $b = -0.064$, $z = -1.326$, $P = 0.185$, 95% CI -0.158 to 0.030.

Second, we see that differences in completion rates based on "WEIRD" country classification hold no matter the chosen motivation, as can be observed from the lower position of the "WEIRD" line (in green) in the middle panel of Fig. 9 across all three facets. Regardless of whether workers work to pay their bills or to kill time, "WEIRD" participants are less likely to work below the contractual minimum. The differences in baseline motivation are therefore not reducible to differences in self-reported reasons for working, and the linear term for the "WEIRD" classification in a regression on the likelihood of completing the entire task remained significant after we controlled for workers' self-reported motives, $b = -0.826$, $z = -4.021$, $P < 0.001$, 95% CI -1.282 to 0.442 (Appendix Table S26).

Continuous Measures of Country-Level Differences

The "WEIRD" classification allows for simplicity in conceptualization and visualization at the expense of erasing many important cultural distinctions. In Additional Table 1 in the main text, we report correlations between proportion completed and a number of continuous country-level measures, such as collectivism, rule of law, and GDP. The correlation with GDP per capita (r = 0.81) is shown in Additional Fig. 10, and similar figures for responsibilism (Talhelm et al., under review) collectivism (Pelham et al., 2022), rule of law (World Justice Project, 2023) are presented in Appendix Figures S15a, S15b, and S15c.

Incomplete Contracts and Ambiguous Pay in Study 5B

The Appendix section "Study 5B" and Figures S6, S7a, and S7b report results from 7,578 additional participants who completed the same task with a more relational framing, where we told workers that "their pay will be finalized after the task," making the lack of additional money for going above and beyond the contractual requirements ambiguous. We found that the difference in effort we saw in Study 5a held when extra pay was implied but not explicitly promised. We also found that the difference in completion rates between "WEIRD" and "non-WEIRD" countries was significantly lower in this ambiguous condition in Study 5B compared to the "Explicitly No Pay" condition in Study 5A, which suggests that contractual ambiguity with regard to additional pay might have a larger motivating effect in Western countries.

DISCUSSION

How effective are different incentives in motivating effort? The data here suggest that the answer depends on culture. In Studies 2a–c, the motivating power of money over psychological interventions was stronger in the US and UK than in China, India, Mexico, and South Africa. In China and Mexico, a social norm intervention was more cost-effective than a monetary

incentive. In Study 4, we found that randomly assigning people in India to take the study in English (compared to in Hindi) increased the money advantage. Participants who took the study in English were also more likely to report that they completed the task "only for money," which suggests that they thought of their work more in accordance with the transactional norms of market exchange.

Finally, in Study 5, we observed that crowd-source workers in forty countries differ in their propensity to work beyond the minimum required to receive pay. This suggests that the money advantage might vary across countries at least in part because workers in "non-WEIRD" countries are more motivated to put in effort above the contractual minimum without explicit financial rewards.

We interpret our results as reflecting cultural differences in exchange norms and psychological contracts. However, one can argue that our findings show that people in "WEIRD" cultures are closer to the model of Homo Economicus, particularly in its narrow reading that includes the "selfishness axiom" (Thaler, 2000), or the assumption that "individuals seek to maximize their own material gains…and expect others to do the same" (Henrich et al., 2005). If maximizing material gains is the sole goal, our participants should work only when it pays. Many Americans did just that, living up to Benjamin Franklin's dictum that "time is money" (Franklin, 1748).

Over half of the participants from the US in Study 2b quit the task as soon as they could do so without losing the base pay when there was no monetary incentive to continue. In Mexico, over 90% of participants continued even when they knew that extra effort would not result in more pay (see section "Probability of Quitting the Task at the First Opportunity in Studies 2a–4" of the Appendix). In Study 5a, 78% of workers in "WEIRD" countries did not complete the

entire data entry task when doing so would not bring them extra pay. In "non-WEIRD" countries, the majority of the workers (54%) did.

We do not see the Homo Economicus interpretation as contradictory to our focus on transactional contracts and exchange norms. Both make the same prediction regarding effort in the non-monetary conditions (H. Chen et al., 2018; Ferraro et al., 2005). However, Homo Economicus has become somewhat of a strawman (Binmore, 2005), particularly since most economists agree that rational agents can derive utility not solely from material gains, but also from moral and reputational considerations (Boyer & Petersen, 2018; Loewenstein et al., 1989).

Both working above the contractual minimum and stopping at 10 images or 5 minutes when there is no monetary incentive to continue can be seen as rational utility-maximizing actions, and therefore we do not interpret the behavior of "non-WEIRD" participants in our studies as "irrational." Nor do we evaluate the behavior of "WEIRD" participants as demonstrating an obsession with money. Instead, we theorize that people across cultures differ in how likely they are to see an activity as an exchange of time and effort in return for explicit remuneration (Daniels, 1987; Zaki et al., 2021).

The Definition of Work, and Conceptualizing an Activity as "Work" Versus "Non-Work"

King Solomon's usage of the word "work" cited in the introduction is likely different from its predominant contemporary meaning. The primary definition of "work" in modern English is "to perform work or fulfill duties regularly *for wages or salary*" (Merriam-Webster, 2024, emphasis added). This definition aligns with a central tenet of neoclassical economics that equates "work" with disutility and views monetary compensation as a prerequisite incentive to counterbalance the disutility derived from labor (Jevons, 1871). However, it would be inaccurate

to say that "work" has more positive implications in languages other than English: for instance, in most Slavic languages, such as Russian, Serbo-Croatian, or Ukrainian, the etymology of the word "work" stems from the words "servitude" or "slavery" and used to refer to forced labor performed by slaves or serfs (Vasmer, 1950).

The meaning of "work," both linguistically and more broadly, has changed dramatically since the Industrial Revolution. The rise of capitalism has alienated workers from the means of production and the results of their labor, transforming work into a commodity that is bought and sold in markets (Lukács, 1971; Marx, 1867; Simmel, 1907). Employers generally have the upper hand in the wage-setting process and typically pay workers the minimum amount necessary to bring them to work a particular amount of time or produce a certain amount of output. Workers, often due to economic necessity, are willing to accept lower wages to secure employment. This is especially true for jobs with high labor supply but low skill requirements, such as image classification and data entry—the tasks used in our studies.

Yet, "work" remains a polysemic word across languages (Godelier, 1980). In some languages, such as English, the meaning of the word "work" is broader and can refer to a number of distinct concepts, such as a productive activity ("it takes work to learn coding"), a result of one's labor ("her life's work"), and a place where paid labor is conducted ("I live an hour away from work"). It is distinguished from the concept of "labor," which denotes a concept of unpleasant, often exploited, work for the profit of others (Friedmann, 1960; Fuchs & Sevignani, 2013; Komlosy et al., 2018; Marx, 1867). As W. H. Auden put it, "Between labor and play stands work. Whether a job is to be classified as labor or work depends, not on the job itself, but on the tastes of the individual who undertakes it" (Auden, 1941).

Evidence using word embeddings has shown that the associations English speakers have with the concept of "work" have shifted in the 20th century toward seeing work as a means to an end in form of monetary payment and other extrinsic, material incentives (Walo, 2023). For example, the use of the word "work" next to words such as "accomplished" or "achieved" has declined, but its use around concepts like "earned" has increased. At the same time, a white-collar employee has become an increasingly prototypical image of a "worker" (Kozlowski et al., 2019).

Thus, a possibility complementary to our explanation based on exchange norms and psychological contracts comes from research that showed that categorizing an activity as "work," as compared to "leisure" or "fun," shifts the motivational focus towards more extrinsic goals like earning money (Etkin & Memmi, 2021; Fishbach & Choi, 2012; Glynn, 1994; Laran & Janiszewski, 2011; Tang & Baumeister, 1984). It is possible that Western participants in our studies were more likely to see the task as "work." This could have made them adopt the rules of market exchange and thus prioritize monetary rewards.

If this is the case, our findings are particularly counterintuitive: why would non-Western workers be less likely to view the tasks we gave them as "work?" In 14 countries in our sample in Study 5 (Kenya, Algeria, Zambia, Egypt, Ghana, Uganda, Sri Lanka, Venezuela, Ethiopia, Madagascar), workers report earning less than $5,000 a year, and "non-WEIRD" workers on crowdsourcing platforms—in our data and in other studies (Ipeirotis, 2010; Jiang et al., 2015; Rogstadius et al., 2011)—report relying on crowdsourcing income more than their "WEIRD" peers (Fig. S13).

Limitations

We have no doubt that the features of our experiment design, such as the domain, task, and social context were pivotal for determining the effects of incentives on motivation that we observed. The domain used in our studies was that of work—insofar as work refers to a paid activity that people do for the benefit of another person. The task, instituted on an online crowdsourcing platform—or more accurately, a set of platforms with varying "laws" and "institutional frameworks"—was manual, non-collaborative, and intrinsically meaningless to workers. The workers completed the task alone and have no personal relationship with their employer or colleagues. The wages they receive were low, but all workers received a base salary, meaning that no one worked solely for commission. We also randomly assigned incentives to workers, but some workers in the real world can choose between jobs based on how much those jobs pay and how much fulfillment they give (Wilmers & Zhang, 2022). Because of these unique features of the design, we can only theorize whether the same cultural differences in the money advantage would arise on different tasks and especially outside of the domain of work, and low-wage work specifically.

Moreover, while our experiments focused on the effectiveness of incentives in terms of motivating effort, it is not the sole metric relevant to evaluating their overall value (Bubonya et al., 2017). For instance, working for piece-rates can make people feel more stressed (Baktash et al., 2022) and decrease the amount of time they spend with loved ones, reducing the quality of non-work relationships (Hur et al., 2021). In future research, it is crucial to assess various incentives on other factors, such as people's sense of overall well-being.

Implications

The findings in this paper are a reminder that we should avoid extrapolating the conclusions from studies based on the 12% of the world population that lives in "WEIRD" cultures to the remaining 88% (Henrich et al., 2010). Instead, monetary—relative to

psychological—incentives may increase effort less in "non-WEIRD" cultures. Accordingly, when designing interventions and reward schemes, at least for low-wage work and particularly under limited resources, the relative benefit of adding a pay-per-effort incentive might be attenuated in "non-WEIRD" cultures.

The implication of our studies is that the importance of money in motivating workers in low-wage jobs might be especially pronounced in high-income, Western economies. This contradicts the increased emphasis in Western countries on non-monetary rewards to attract the best employees (Cassar & Meier, 2018; Thibault Landry et al., 2017) and suggests that this higher focus on non-monetary aspects of the job might not generalize to the entire workforce. For workers in low-skilled jobs, pay likely remains the predominant factor that determines job satisfaction.

Ethical Considerations

Our studies speak to several thorny issues about compensation and global inequality. People could use our data to justify paying people in "non-WEIRD" cultures less: the fact that most non-Western workers in our studies worked more than was required of them to receive pay they raises concerns about exploitation.

First, in our design, everyone received a base pay (a "salary"), which was probably the first-order motivator. On crowd-sourcing sites, participants would likely not start tasks in the first place if they were not paid (Ipeirotis, 2010). Thus, we interpret our results as suggesting that, given an adequate salary, workers from "non-WEIRD" cultures are less incentivized by further pay-for-effort incentives as compared to free psychological motivators.

Second, automation, offshoring, and migration to high-income economies have caused growing job polarization in nearly all developed countries. This means that the demand for high-skilled and low-skilled jobs increases at the same time as the demand for middle-skilled jobs

decreases. Around half of the US labor force works low-paying jobs—the highest share among developed countries (Appelbaum & Schmitt, 2009)—earning a median wage of $16 an hour or less. In several large US cities, like Los Angeles and Miami, low-skilled workers make up the majority of the workforce (Ross & Bateman, 2019).

At the same time, developed countries are increasingly relying on migrants to fill labor shortages in predominantly low-wage sectors. The United Nations estimated that there were 169 million international migrant workers, 113.9 million (67.4%) of whom are in high-income countries (International Labour Organization, 2021b). In these settings, migrants often find themselves in precarious work conditions, facing challenges such as wage theft and limited access to legal and social protections (Moyce & Schenker, 2018).

We would argue that crowd-source workers in our studies are likely not comparable to the new Google recruits in Silicon Valley. But they might resemble the 75 million workers employed in "fast fashion" worldwide or a billion people working in agriculture, of which 22 million are in the US (International Labour Organization, 2021a). The majority of them still work for nothing but tiny piece-rates.

Should Behavioral Scientists Intervene?

If our results generalize to other industries, they suggest that non-Western workers often keep working even if they are not being paid more for extra effort. Should behavioral scientists and policy-makers intervene to change the way non-Western workers view paid labor?

We would argue that increasing the market mentality of non-Western workers might not be an appropriate conclusion. Interventions attempting to do so might end up harming communities, as the institutional and social realities in which non-Western workers operate might be different from those of Western workers.

Workers in developing countries (and migrants in developed countries) often face high unemployment together with weak legal protections, bargaining power, and job security (Butcher

& Schanzenbach, 2018; Freeman, 2010; Rama, 2003). They might be more likely to face questions such as, “My employer has to let go of 10 workers next week; how do I make sure it’s not going to be me?” or “How can I increase my chances of being rehired for the next harvest?” This instability can lead to increased effort even when they cannot reap immediate rewards for it, and in the long run, trying harder might pay off. On the other hand, doing just what the contract says might not be “enough” to survive. Instead, we would argue that interventions that increase public awareness and support for structural changes are more appropriate (Madan et al., 2023).

In 2021, California passed the Garment Worker Protection Act and became the first US state to prohibit piece-rate-only payment schemes (without a salary) for garment workers (California Department of Industrial Relations, 2021). These workers, many of whom are immigrant women, must now receive the state’s minimum hourly wage, just like all other workers who do not work for piece-rates. Before, some of these garment workers received piece-rates so minuscule that their wages amounted to as low as $1.58 per hour. This suggests that the tendency of some (often poorer, non-Western) workers to work more for less money is already being utilized in the globalized world. We can only hope for more structural changes that can address the pattern of work effort we see in our studies and ensure that people around the world can earn a fair wage.

METHODS

All studies were carried out in accordance with all the ethical regulations and were approved by the Institutional Review Board (IRB15-1623 and IRB20-1056) at the University of Chicago. Informed consent was obtained from study participants consistent with the IRB protocol.

Methods for Study 1

Categorizing Treatments

The original DellaVigna and Pope (2018) experiment included a total of 18 incentive treatments, to which participants from both India and the US were randomly assigned. Out of those 18 treatments, we first selected all pay-for-effort treatments with linear, immediate, and guaranteed piece-rates (e.g., "You will be paid an extra 1 cent for every 100 points") and categorized them as "monetary." These monetary treatments included a range of pay-for-effort incentives with varying linear piece-rates, ranging from 1 cent per 1,000 button-presses to 10 cents per 100 button-presses.

Using the criteria of linear, immediate, and guaranteed piece-rates, we excluded those conditions that, for instance, tested for loss aversion ("You will be paid an extra 40 cents if you score at least 2,000 points") or delay discounting ("You will be paid an extra 1 cent for every 100 points" with a four-week delay). The reasons for these exclusions were that the former conditions only offered additional (lump) payment upon reaching a high number of presses, while the latter offered extra payment that was not immediate.

To create a list of non-monetary conditions, we selected treatments where the participants could not earn any additional money for themselves, neither immediately nor at some point in the future, and categorized them as "non-monetary." The non-monetary treatments included

conditions with incentives labeled by DellaVigna and Pope (2018) as "social psychological" (such as the social norm treatment ("Many participants scored more than 2,000"), the two charity conditions, where the participants could earn money for the Red Cross but not for themselves, as well as the flat fee "control" condition ("Your score will not affect your payment").

Appendix Table S27a summarizes all the treatments we included based on these criteria and Appendix Table S27b lists all the treatments we excluded, together with the corresponding reasons for exclusion. Our final sample consisted of 6,294 participants: 5,526 in the US and 768 in India. Demographic information is available in the Appendix Table S29a.

Methods for Study 2a–c

Establishing the Minimal Amount of Effort

Instead of the button-pressing task, we asked participants to classify whether images contained a building or not. We kept the original ten-minute limit but made it explicit to participants (including a comprehension check) that they were free to quit the task after every 10 image-ratings without losing their base pay. We did this to remove any ambiguity regarding whether participants would be punished for not exerting a lot of effort, particularly in the non-monetary conditions, and because we were worried that "non-WEIRD" participants might be more inclined to think they might be penalized for doing so.

Internet Connectivity

We also assumed that people might be faster at rating images on their computers than on their phones or tablets and that people in different countries might have different Internet speeds. While the main effect of culture on effort was not our primary variable of interest, we still chose to limit the participants to laptop or desktop computers only. We also asked participants to self-report how long it took for images to load. This allowed us to rule out technological differences

across countries as an explanation for the pattern of findings. Analyses controlling for Internet connectivity are presented in the Appendix Table S31a.

Work Meaning

Because the meaning behind work matters for the effectiveness of incentives (Ariely et al., 2008; Kosfeld et al., 2017), we wanted to replicate the findings from Study 1—in which the work was explicitly meaningless—on a task where participants would be provided with some purpose behind the monotonous work. Therefore, we told participants that image-classifications would help the researchers with "training a machine-learning algorithm." The main rationale behind this change was the assumption that most real-life work tasks are not as devoid of meaning as pressing two buttons interchangeably for no apparent reason.

Base Pay and Piece-Rates

Our goal was to select a piece-rate that would provide a sufficiently strong incentive for the participants to exert effort. However, we did not want to the piece-rate to be so high as to remove any meaningful variation. We agreed to set the base pay across studies to $1.30, and the piece-rate in the monetary condition to 5 cents per 10 images (unless otherwise noted). Participants did not receive partial pay for rating fewer than 10 images within each increment. That is, a participant who rated 11 images and a participant who rated 19 images would both receive a bonus of 5 cents.

Based on a pilot, we expected the participants in the monetary condition to earn, on average, an additional 30 to 50 cents, a nominal amount and proportion of final earnings (relative to base pay for taking the study) between those of the "1 cent per 100 presses" and "4 cents per 100 presses" conditions in DellaVigna and Pope (2018). However, we estimated that participants

working their hardest would be able to complete at least 150 images within the span of 10 minutes and thus earn 75 cents.

Incentive Treatments

We wanted to have the minimal number of conditions to meaningfully test our hypothesis with a sufficiently large sample. Across Studies 2a–c, we kept one pay-for-effort incentive condition and two non-monetary conditions.

In Studies 2a and 2b, the non-monetary conditions included a social norm condition, where the participants were told that many other participants tried hard on the task and assessed 160 images, and a flat fee condition, where no additional instructions were given. In Study 2c, we changed the non-monetary treatments. In one of them, participants read that the task was a competition and that they would see how they did relative to others upon completion (competition condition). In the other, participants read that they would not receive any additional pay; however, the researchers would transfer 5 cents for every 10 images participants rated to the Red Cross (charity condition).

In all the non-monetary conditions, the amount of money one earned was not contingent on how much effort one exerted. Every participant received the same salary regardless of how many images they rated. Yet, receiving a fixed salary might invoke norms with regards to obligations and reciprocity, which can differ across cultures (Hitokoto, 2016; Shen et al., 2011). Since the pay-off structure was the same across all the non-monetary conditions, we first pooled across the non-monetary conditions in each study and compared them to the monetary condition.

Procedure

Participants first read the consent form. Then they read about the nature of the task and were told that they would receive their base pay if they completed at least 10 image-ratings.

Next, participants had to pass an attention check that asked them about the purpose of the task mentioned on the previous page (to train a machine-learning image-classification database). After that, they learned about the structure of the task and their condition assignment. They had to pass two comprehension checks: one asking them about the maximum duration of the task (10 minutes) and the other asking them whether they would receive additional pay based on how many images they rated. Further detail and exact wordings are provided in section "Attention and Comprehension Checks" of the Appendix.

Only those participants who passed the attention check and both comprehension checks proceeded to the image-rating task that contained the dependent variable. The participants saw several images one by one, with a 10-minute timer visible. Images included pictures of plants, flowers, urban landscapes, and buildings (for examples of images used in the picture-rating task, see Appendix Images S1 and S2). For each image, the participants had to answer whether it contained a building.

After rating the first 10 images, the participants saw a screen that asked them whether they wanted to quit or continue with the task. If they chose to quit, they proceeded with the other questions in the survey. Participants who continued with the image-rating task saw a similar screen after rating every 10 images.

Cost Effectiveness

To calculate the amount of effort per dollar spent, we divided the number of images each participant rated by the pay they received. Confidence intervals for the amount per dollar spent are 95% confidence intervals of the mean effort divided by the cost in dollar amount adjusted for each level of effort. This cost would be the same for the non-monetary conditions for all amounts of effort within the confidence interval (except for the charity condition, which entailed an

additional cost for the researcher). However, it would vary in the monetary conditions (i.e., would be higher for higher values within the confidence interval).

Study 2a

Participants

In China, we recruited participants (N = 1,086) who were either students at Hubei University (N = 188) or users of two major social media in China: WeChat Moments (N = 79) or QQ Zones (N = 819). In the UK, we recruited participants on Prolific (N = 1,067). A colleague from Hubei University completed the translation. Demographic information is available in the Appendix Table S29b.

Incentive Treatments

Participants were randomly assigned to one of the three conditions: monetary, flat fee, and norm. The base pay was ¥8.25 ($1.30 at the time of the experiment) in China and $1.30 in the UK. In the monetary condition, the bonus for every additional 10 images in the monetary condition was ¥0.30 ($0.05) in China and $0.05 in the UK.

Methods for Study 2b

Participants

Participants for all three samples were recruited from Prolific: N = 1,053 in Mexico; N = 1,098 in the US sample with the same nominal pay; N = 1,122 in the US sample with increased pay pre-tested to be subjectively equivalent (Becker et al., 1964) to that in Mexico. The participants in Mexico completed the study in Spanish, while those in the US did so in English. A professional translator and a Spanish-native research assistant completed the translation. Demographic information is available in the Appendix Table S29b.

Incentive Treatments

Participants were randomly assigned to one of the three conditions: monetary, flat fee, and social norm. The incentive treatments were identical to those in Study 2a, except for the increase in pay in one of the US samples, as described below. Participants in Mexico received a $1.30 base pay for completing the study and, in the monetary condition, a $0.05 bonus per 10 image-ratings. The two US samples (described in the next section) differed in both the flat fee that the participants received ($1.30 versus $1.56) and in the piece-rate ($0.05 versus $0.06 per 10 images) in the monetary condition.

Becker-DeGroot-Marshak Procedure

To address the limitation of Study 1 and 2a—namely, the potentially different subjective values of the pay amounts between the samples—for Study 2b, we collected three samples of participants on Prolific: one from Mexico (N = 1,053) and two from the US, one with the same nominal pay as in Mexico (N = 1,098) and the other with the same subjective pay as in Mexico (N = 1,122). Participants in the US completed the study in English, while those in Mexico did so in Spanish.

To determine subjectively equivalent pay amounts for Prolific participants from Mexico and from the US, we followed the Becker-DeGroot-Marschak (BDM) procedure (Becker et al., 1964). A separate group of participants got briefly acquainted with the task and were then asked how much remuneration they would need to complete its full version. A detailed description of the procedure and the results is provided in section “Becker-Degroot-Marshak (BDM) Procedure for Establishing Pay Equivalence” of the Appendix.

Methods for Study 2c

Pre-Registration

The study was pre-registered on February 3, 2023. The pre-registration is available on AsPredicted: https://aspredicted.org/dm562.pdf. The study did not deviate from the pre-registration. We pre-registered a total of 665 participants per to have enough power to detect an effect size of $f = 0.14$ for the interaction between incentive type and culture.

Participants

As in Study 2b, we recruited three samples of participants on Prolific: one from South Africa ($N = 649$) and two from the US, one with nominally the same pay as in South Africa ($N = 662$) and the other with subjectively the same pay as in South Africa ($N = 662$). All participants completed the survey in English. We pre-screened participants in South Africa who were fluent in English. Demographic information is available in the Appendix Table S29b.

Incentive Treatments

Unlike in Studies 2a and 2b, participants in this study were first randomly assigned to types of incentives (monetary or non-monetary) and not individual incentive treatments. Those assigned to the non-monetary incentive type were then randomly assigned to either the competition or the charity condition. Thus, incentive types have roughly identical numbers of people per cell (~330), but each individual non-monetary treatment had half as many people (~165) as the monetary condition. In our previous studies, the number of participants was the same in each individual condition, not in each type of incentive condition (monetary or non-monetary). The pay structure in the monetary condition is described in the next section.

Becker-DeGroot-Marshak Procedure

To determine subjectively equivalent pay amounts between Prolific participants from South Africa and from the US, we similarly followed the Becker-DeGroot-Marschak (BDM) procedure (Becker et al., 1964). A separate group of participants got briefly acquainted with the

task and were then asked how much remuneration they would need to complete the full version of the survey.

The two US samples differed in both the flat fee that participants received ($1.30 versus $2.25) for participation and in the pay-for-effort rate ($0.05 versus $0.09 per 10 images) in the monetary incentive condition. Participants in South Africa received a $1.30 base pay for completing the study in all three conditions and a $0.05 bonus per 10 image-ratings in the monetary incentive condition. A detailed description of the procedure and the results are available in section "Becker-Degroot-Marshak (BDM) Procedure for Establishing Pay Equivalence" of the Appendix.

Methods for Study 3a

Pre-Registration

The study was pre-registered on August 25, 2022, though we deviated from the pre-registration as described below. The pre-registration is available on AsPredicted: https://aspredicted.org/uz5gh.pdf.

The main deviation concerned the analysis plan. We pre-registered data analysis from both countries separately using *t*-tests, with a prediction that, for people in the US, a small monetary incentive would result in higher effort than emphasizing the social norm; our second prediction was that the two incentives would be statistically indistinguishable from each other in India. While these hypotheses were supported by the data, we believe that we did not have sufficient power to detect the smallest meaningful effect size in India and thus the pre-registered analysis was inappropriate. Instead, we ran a multiple linear regression model, consistent with other studies in the paper, to probe for the interaction between incentive and country. We present

the results of the pre-registered analyses in section "Additional Analyses for Study 3A" of the Appendix.

Second, we recruited participants in the US on Prolific instead of MTurk. The pre-registration stated that recruitment would take place on MTurk. We made this change because we noticed that some US MTurkers had posted reviews of the previous reiterations of the study informing other MTurkers that it was sufficient to complete only 10 images to receive the pay, which we thought might compromise the quality of the data.

Third, we pre-registered 352 participants per country after exclusions (i.e., people who did not rate a single image or those who did not complete the full study and receive payment) to have 80% power to observe a small to medium-sized effect (d = 0.30). We sampled more people to allow for exclusions and repeated submissions. (Exclusion details for all studies are available in section "Exclusion Data and Criteria for Studies 2–5" of the Appendix.) Fewer people than we had estimated did not meet the inclusion criteria in the US, and so the final sample included 382 people in the US and 352 people in India. Excluding the last 30 participants in the US did not significantly change any of the main results, and these analyses are reported in section "Additional Analyses for Study 3A" of the Appendix.

Participants

In India (N = 352), we recruited people on MTurk. In the US (N = 382), we recruited people on Prolific. Participants in both countries completed the survey in English. Demographic information is available in the Appendix Table S29c.

Incentive Treatments

Participants were randomly assigned to one of the three conditions: minimal pay or norm. Procedure was identical to the one used in Studies 2a–2c, except for the base pay that was $1.00

(as opposed to $1.30 in previous studies) in both India and the US across conditions. The norm condition was identical to the one in Studies 2a–2b. In the minimal-pay condition, the participants could earn an extra cent for completing every 20 image-ratings.

Methods for Study 3b

Pre-Registration

The study was pre-registered on March 29, 2023. The pre-registration is available on AsPredicted: https://aspredicted.org/7e2xy.pdf. We chose not to pre-register a specific hypothesis and instead delineated how the results would be interpreted based on each possible outcome of the analysis. The study did not deviate from the pre-registration. We pre-registered recruiting a total of 540 participants in the US to have 90% power to detect an effect size of f = 0.14.

Participants

We recruited the participants in the US on Prolific (N = 537). Demographic information is available in Appendix Table S29c.

Incentive Treatments

Participants were randomly assigned to one of the two conditions: minimal pay or points (gamification). In both conditions, participants received $1.30 for participation. In the minimal pay condition, participants read that they would receive an additional 1 cent for every 10 pictures they rated. In the points (gamification) condition, participants were told that they would receive no additional pay but would get a point for every 10 pictures they assessed. The instructions further explained: “You can treat these points as if it were a game. However, they do not lead to additional financial compensation.”

We raised the pay from 1 cent for every 20 images in Study 3a to allow participants to receive a cent (or a point) on every screen that asked them whether they wanted to continue with or quit the task (after rating every 10 images), which we thought would be more conducive to gamification.

Methods for Study 4

Participants

We recruited participants through the Center for Social and Behavior Change at Ashoka University, a private university in Haryana, India. The final sample consisted of 2,065 participants recruited through advertisements on the Facebook group "QMaths." This group has over 280,000 members interested in preparing for competitive exams for jobs in sectors ranging from banking to railways. We selected this Facebook group because 1) Ashoka had an ongoing relationship with one of the moderators of the group and 2) the members would generally be proficient in English and Hindi. Demographic information is available in the Appendix Table S29d.

Language

Participants were randomly assigned to complete the survey in Hindi or in English. Participants in the Hindi condition completed the whole survey, starting with the consent form, in Hindi. Participants in the English condition completed the entire survey in English. Two research assistants from Ashoka University completed the translation from English into Hindi, occasionally changing the original English wordings to ensure compatibility between the two languages.

We included checks to ensure that participants were proficient in both languages. One week before taking the survey for Study 4, participants completed another survey for which they had to report being "very good" or "fluent" speakers of both English and Hindi.

Incentive Treatments

Participants were randomly assigned to incentive and language conditions. To ensure a sufficiently high number of participants per condition, we kept two conditions: one monetary condition and one non-monetary condition (social norm). Therefore, the study followed a 2 Language (English, Hindi) x 2 Incentive (Monetary, Non-Monetary) between-subjects design. The social norm condition was the same as in the previous studies. In the monetary condition, participants received a monetary bonus of ₹5 rupees ($0.0665) for every 10 images they rated. Everyone received ₹150 ($1.995) for completing the main experimental task. To calculate cost-efficiency, we used the exchange rate on the day the last response was collected (December 7, 2021; ₹1 = $0.0133).

Additional Variables

We included seven exploratory variables to measure participants' perceptions of the task and motivations for completing it. These variables were all measured on a 7-point Likert scale (1 = "Strongly Disagree" to 7 = "Strongly Agree"). Participants reported their agreement with the following statements: "I enjoyed completing the task"; "I am satisfied with how well I did on the task"; "I believe that I helped others by completing the task"; "Completing the task was boring"; "I only completed the task for money"; "I could have assessed more pictures if I'd tried harder"; *"*I am satisfied with the amount of pay I received." Details are available in section "Exploratory Variables" of the Appendix.

Methods for Studies 5a and 5b

Conditions

The instructions for the Explicitly No Pay condition in Study 5a read: “You have to spend at least 5 minutes transcribing the sheet. After 5 minutes, you will be able to proceed to the survey. It's up to you whether you want to work longer and complete transcribing the sheet. You will not get paid more if you spend more than 5 minutes. Regardless of how much time you spend (as long as it's above 5 minutes) or how much you transcribe, you will be paid exactly the same amount of money.”

The instructions for the Ambiguous Pay condition in Study 5b read: “You have to spend at least 5 minutes transcribing the sheet. After 5 minutes, you will be able to proceed to the survey. It's up to you whether you want to work longer and complete transcribing the sheet. Your pay will be finalized after the task.”

Participants

We recruited participants on seven crowd-sourcing platforms. The composition of sample from each country is available in the Appendix Table S29f and the demographics in Appendix Table S29e. Everyone but workers in Japan and China recruited on local platforms completed the study in English.

Pay Attention Check

The question read, “What did the instructions say about whether you would not or would not receive more pay for spending more than 5 minutes on the data entry task?” The possible answer choices were: 1) “The instructions said that I would receive more pay for spending longer than 5 minutes;” 2) “The instructions said that I would not receive more pay for spending longer than 5 minutes;” 3) “The instructions did not explicitly say whether or not I would receive more pay for spending longer than 5 minutes;” 4) “I do not remember/I am not sure.”

Secondary Measures

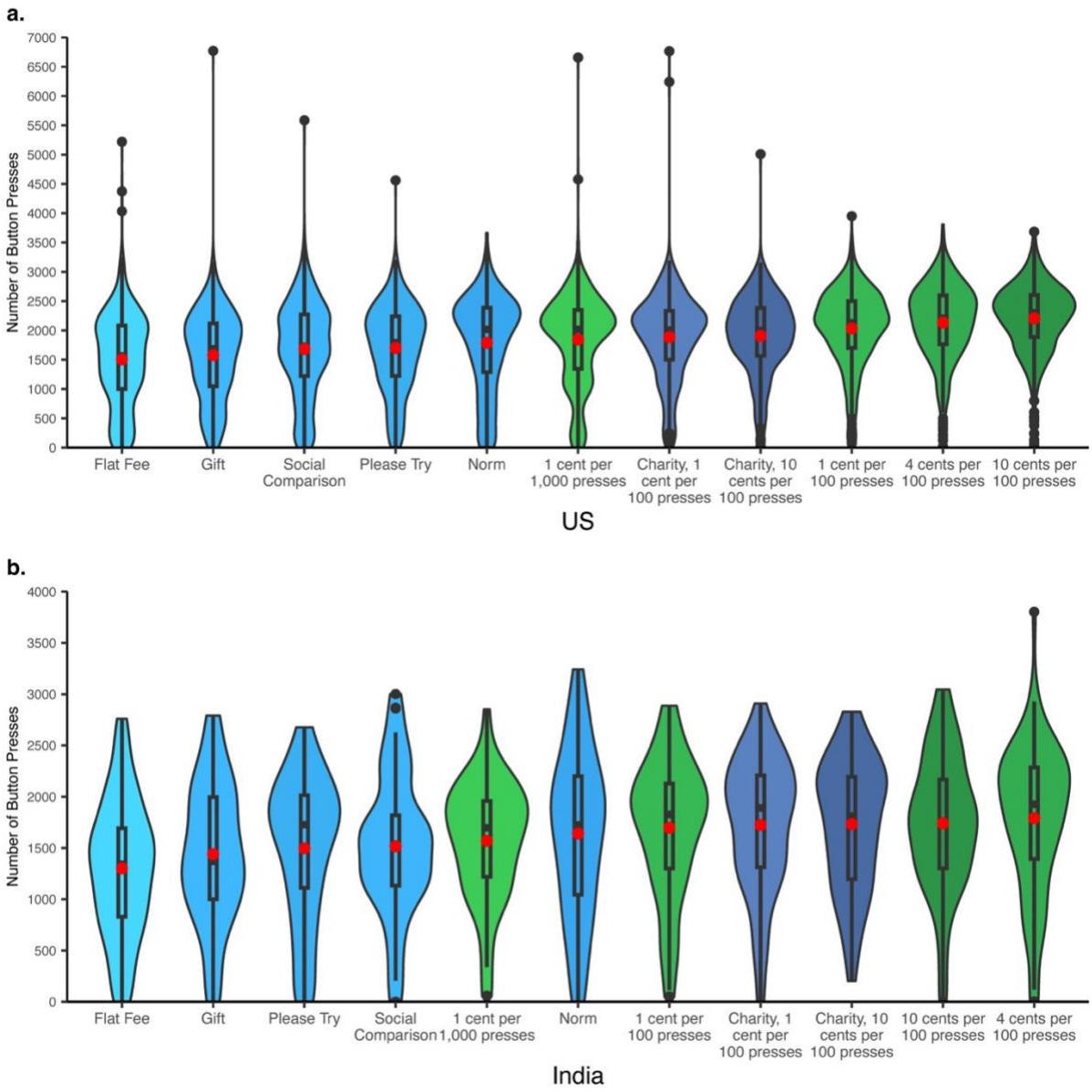
The precise wordings of our secondary measures that tapped into motivations and expectations of future benefits were: “The morally right thing to do was to spend more than 5 minutes on the task;” “I would not trust those who spent only 5 minutes on the task;” “People who didn't complete the whole task are lazy;” “People who didn't complete the whole task are not nice;” “People who didn't complete the whole task are intelligent;” “People who didn't complete the whole task are selfish;” “The researchers wanted me to work for longer than 5 minutes;” “Working for longer than 5 minutes could have earned me more pay;” “I enjoyed completing the task;” “Working for longer than 5 minutes could encourage the researchers to give me more work in the future.”

Social Evaluations Index Reliability

The Cronbach's alpha (α) for the index was 0.77, indicating acceptable internal reliability. Reliability would increase to 0.85 if the intelligence measure is dropped, perhaps suggesting that the other items have more to do with the warmth dimension.

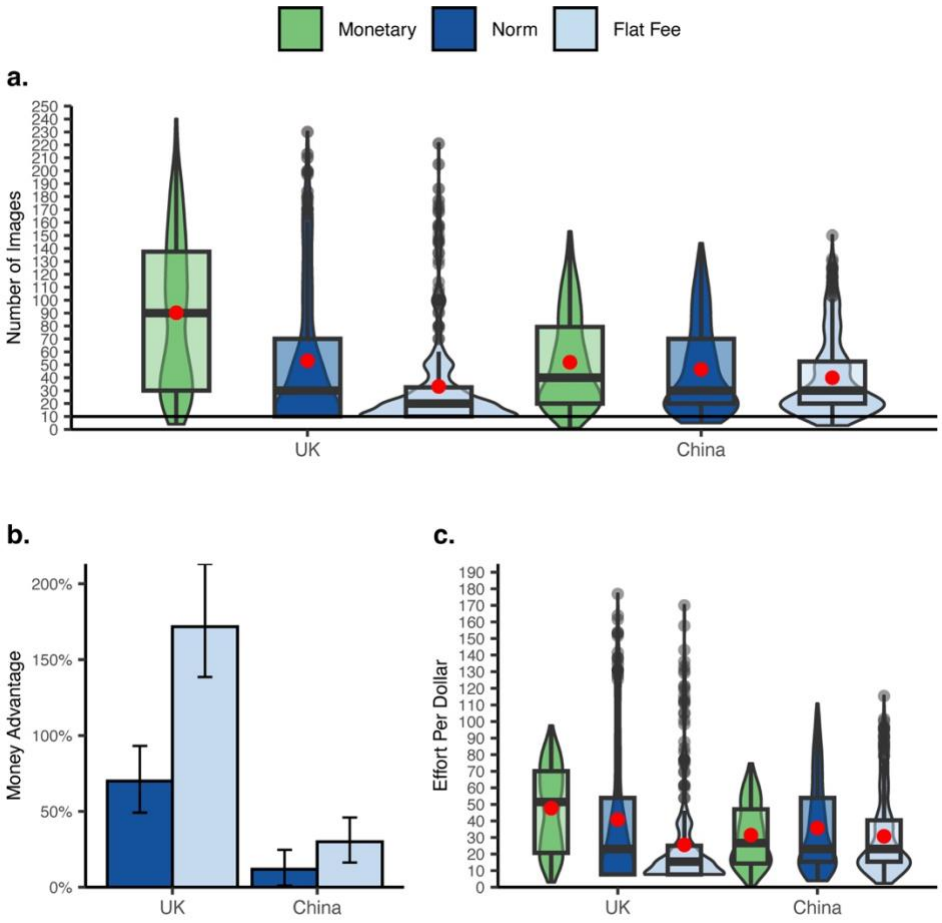
ADDITIONAL FIGURES

Additional Fig. 1: Study 1, Individual Conditions by Country.



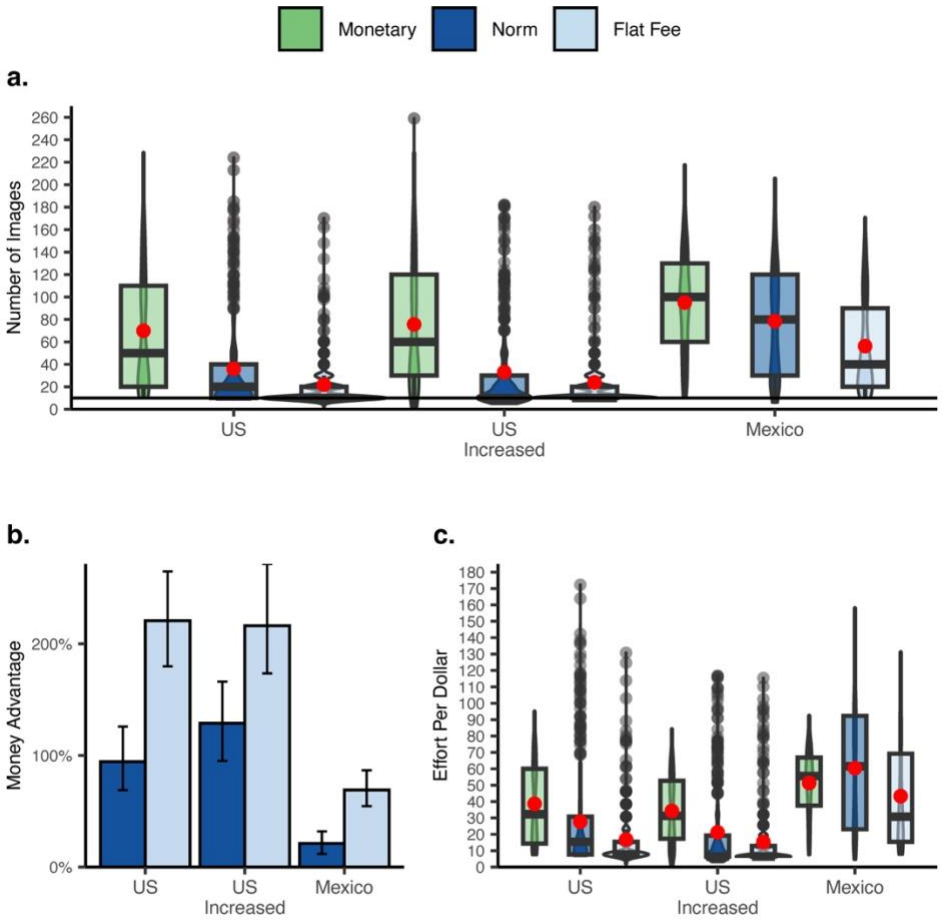
Effects of individual monetary incentives (green) and non-monetary treatments (blue) amongst participants in the US (N = 5,526 on MTurk) and India (N = 768 on MTurk) from a prior study (DellaVigna & Pope, 2018). Panels A (US) and B (India) show the central tendency and distribution of effort by incentive. Conditions within each panel (x-axis) are ordered from the lowest to highest mean effort in each country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations within each panel at any given number of button-presses on the y-axis.

***Additional Fig. 2*: Study 2a, Monetary Versus Individual Non-Monetary Conditions in the UK and China.**



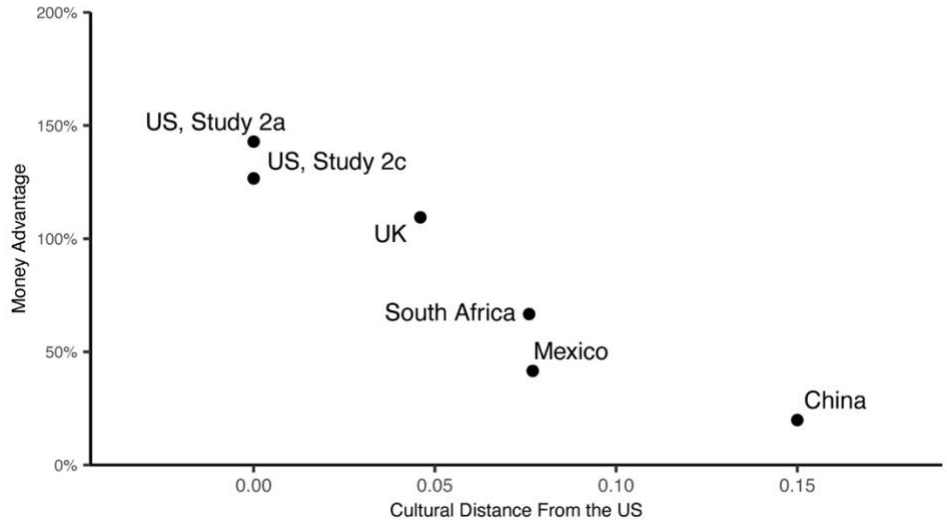
Effects of the three individual conditions: monetary (green), flat fee (light blue), and social norm (dark blue) in the UK (N = 1,067 recruited on Prolific) and China (N = 1,086 recruited on social media). Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. Panel B shows the money advantage—that is, how much more effective money is than each of the two non-monetary treatments in each country. Panel C shows the central tendency and distribution of cost-effectiveness (effort per dollar spent) by incentive type and country. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The results for cost-effectiveness are summarized in the main text. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images in the monetary and each of the non-monetary conditions.

***Additional Fig. 3*: Study 2b, Monetary Versus Individual Non-Monetary Conditions in the US and Mexico.**



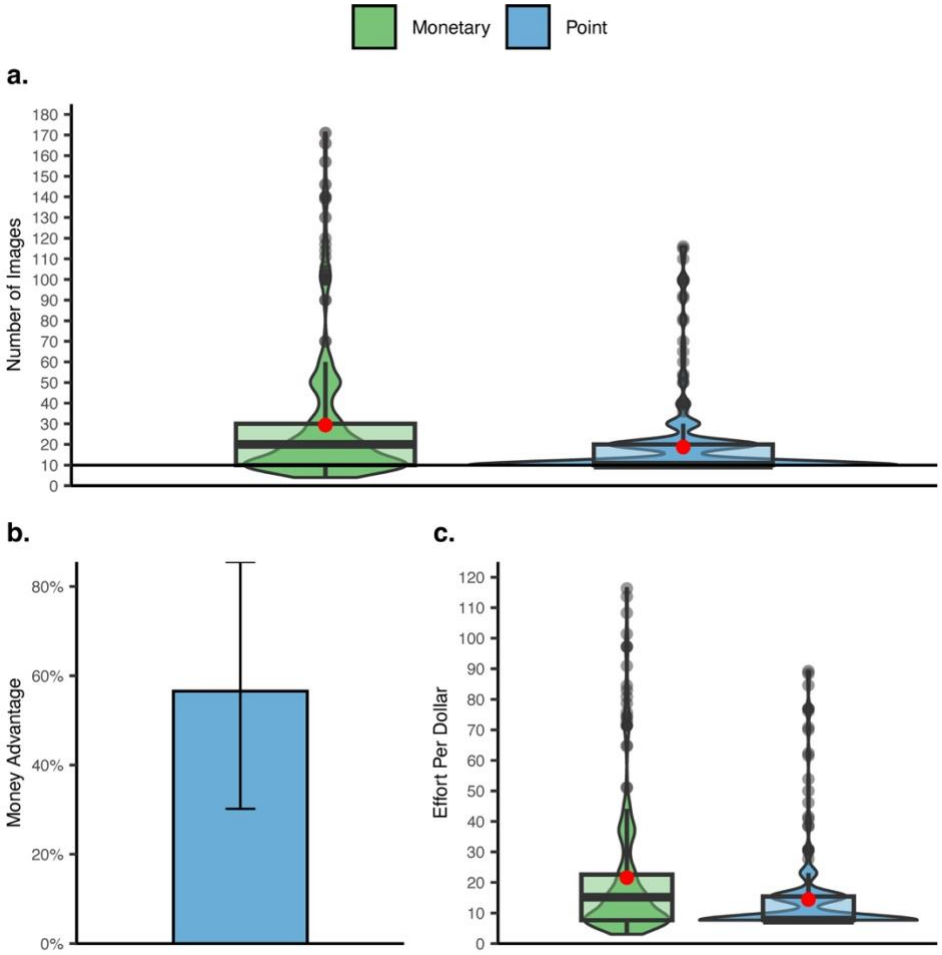
Effects of the three individual conditions: monetary (green), flat fee (light blue), and social norm (dark blue) in Mexico (N = 1,053 recruited on Prolific) and two samples in the US: one with the same nominal pay (N = 1,098 recruited on Prolific) as in Mexico and the other with the same subjective (Becker et al., 1964) pay (N = 1,122 recruited on Prolific) as in Mexico. Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. Panel B shows the money advantage—that is, how much more effective money is than each of the two non-monetary treatments in each sample. Panel C shows the central tendency and distribution of the cost-effectiveness (effort per dollar spent) by incentive type and country. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The results for cost-effectiveness are summarized in the main text. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images in the monetary and each of the non-monetary conditions.

Additional Fig. 4: **Studies 2a–c, Mapping the Money Advantage on Cultural Distance from the US.**



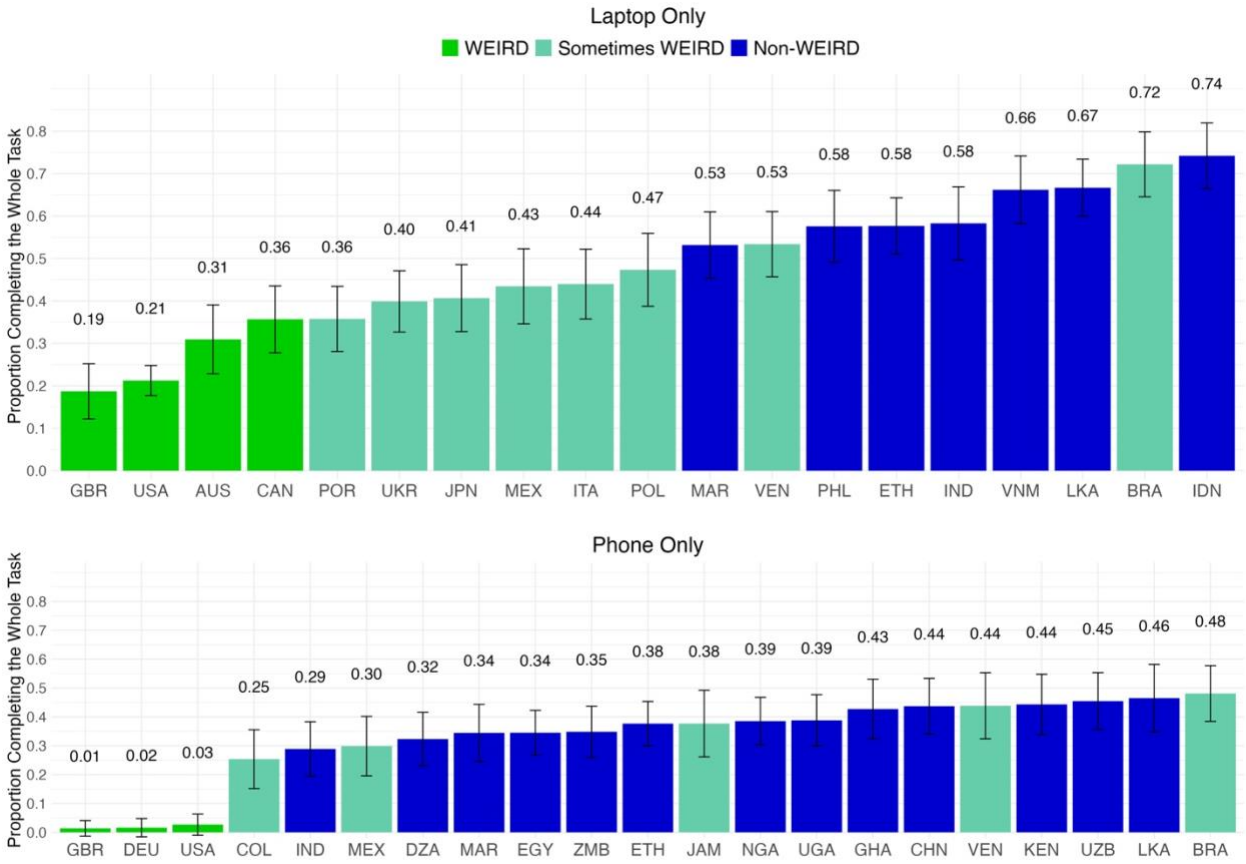
Mapping the difference in the effectiveness of monetary and non-monetary treatments (money advantage) from the samples with the same nominal pay in Studies 2a–c on the cultural distance from the US. Cultural distance scores are taken from Muthukrishna and colleagues (Muthukrishna et al., 2020).

Additional Fig. 5: **Study 3b, Minimal Pay Versus Points (Gamification) Condition in the US.**



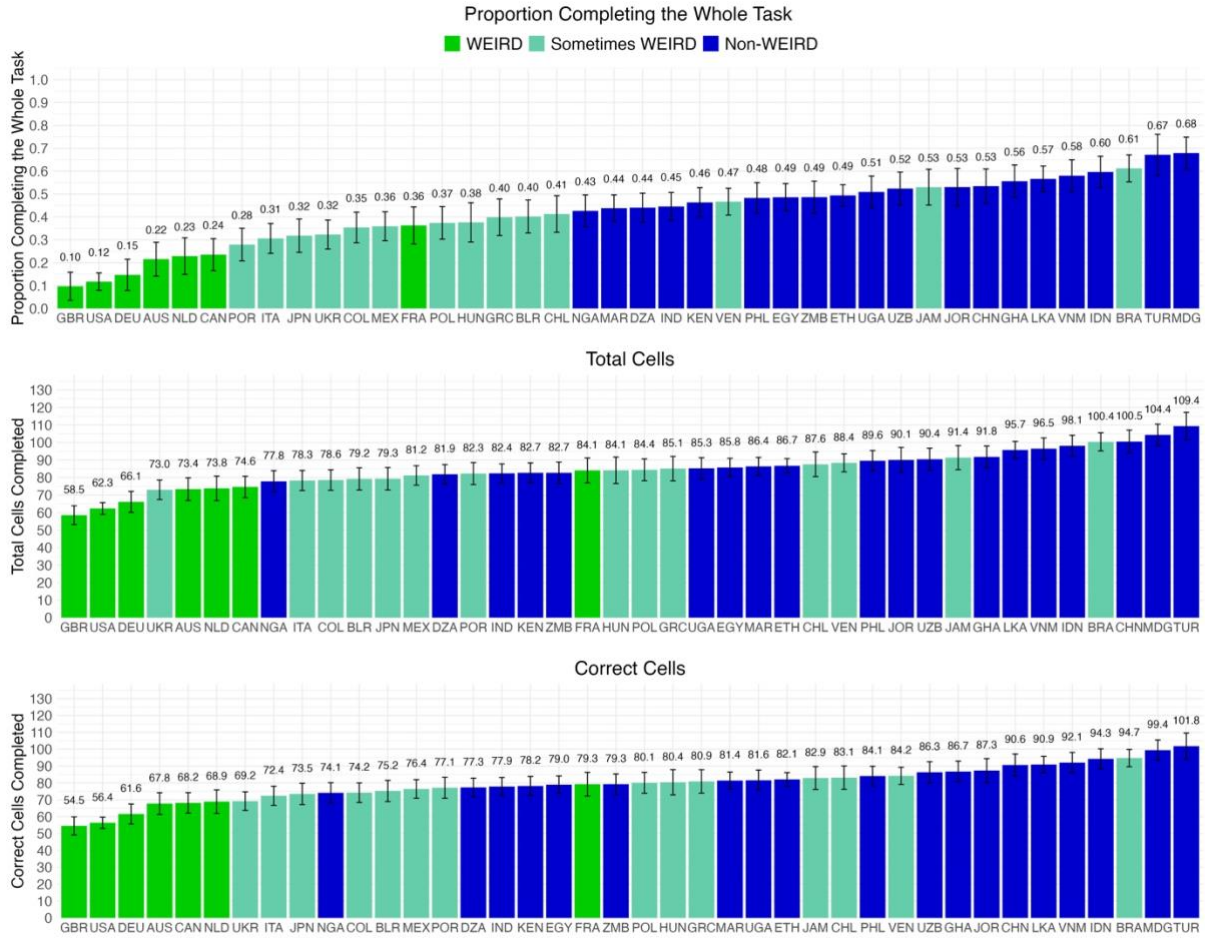
Effects of the minimal monetary incentive (1 cent per rating 10 images; in green) and a non-monetary gamification treatment (1 extra point per rating 10 images; in blue) in the US (N = 537 on Prolific). Panel A shows the central tendency and distribution of effort by incentive condition. The black line within each box represents a median; the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The difference in the number of images rated in the two conditions is statistically significant, Welch's $t(446.98)$ = 4.67, $P < 0.001$, $Mean_{difference}$ = 10.61, d = 0.40, 95% CI 6.15 to 15.08. Panel B shows the money advantage—that is, how much more effective the minimal monetary incentive is compared to the gamification condition. Panel C shows the central tendency and distribution of cost-effectiveness (effort per dollar spent) of each incentive. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The minimal monetary incentive is more cost-effective than the gamification treatment, Welch's $t(471.93)$ = 4.37, $P < 0.001$, $Mean_{difference}$ = 7.15, d = 0.38, 95% CI 3.93 to 10.36. In Panel B, the error bar is a bootstrapped 95% CI for the mean relative difference in the number of images rated in the minimal-monetary-incentive versus social-norm condition.

***Additional Fig. 6*: Completion Rates Among Those Working on Laptops and Phones.**



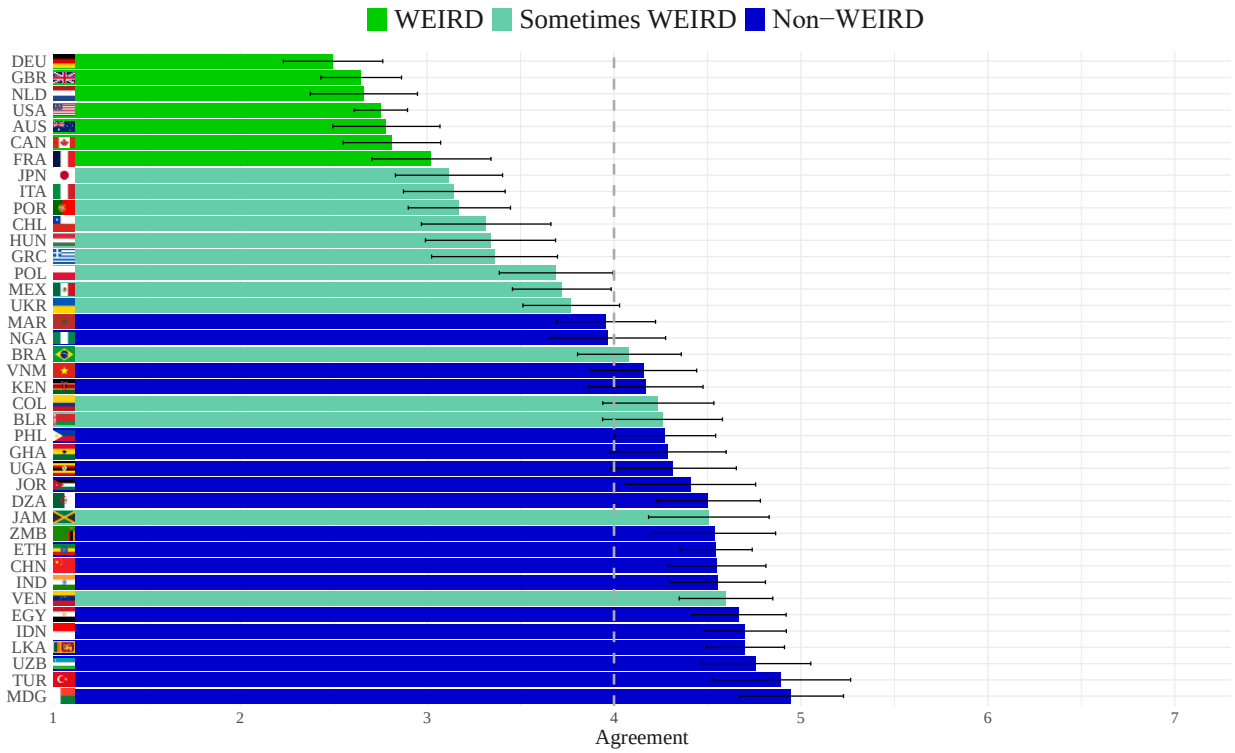
Completion rates in Study 5a (N = 7,823 independent participants recruited on seven crowd-sourcing platforms) separated by whether they completed the task on their laptops (top) or phones (bottom). In the top panel, only countries where at least 120 participants completed the task on their laptop are present; in the bottom figure, we only included countries where over 60 participants completed the task on their phones. The device used was self-reported. The mixed-effects logistic regression analysis showed a significant negative linear effect of the "WEIRD" country classification on the probability of completing the whole task for participants using laptops, b = -1.035, z = -6.070, P < 0.001, 95% CI -1.370 to -0.701, indicating that participants from more "WEIRD" countries who worked on their laptops were less likely to complete the entire transcription. Additionally, the quadratic term for the "WEIRD" classification also showed a significant negative effect for laptop users, b = -0.315, z = -2.494, P = 0.013, 95% CI -0.562 to -0.067, suggesting the decline in completion rates among "WEIRD" countries compared to "non-WEIRD" and "Partly (Sometimes) WEIRD" countries was not linear. For participants using phones, there was also a significant negative linear effect of the "WEIRD" country classification, b = -0.992, z = -5.559, P < 0.001, 95% CI -1.342 to -0.642, similarly indicating lower likelihood of task completion in more "WEIRD" countries. However, the quadratic term for phone users was not significant, b = -0.111, z = -0.957, P = 0.338, 95% CI -0.339 to 0.117, suggesting that the decline from "non-WEIRD" to "WEIRD" countries was not more pronounced among phone users in "WEIRD" or in "Partly (Sometimes) WEIRD" countries. Error bars are 95% CIs.

Additional Fig. 7: Additional Measures of Effort in Study 5A.



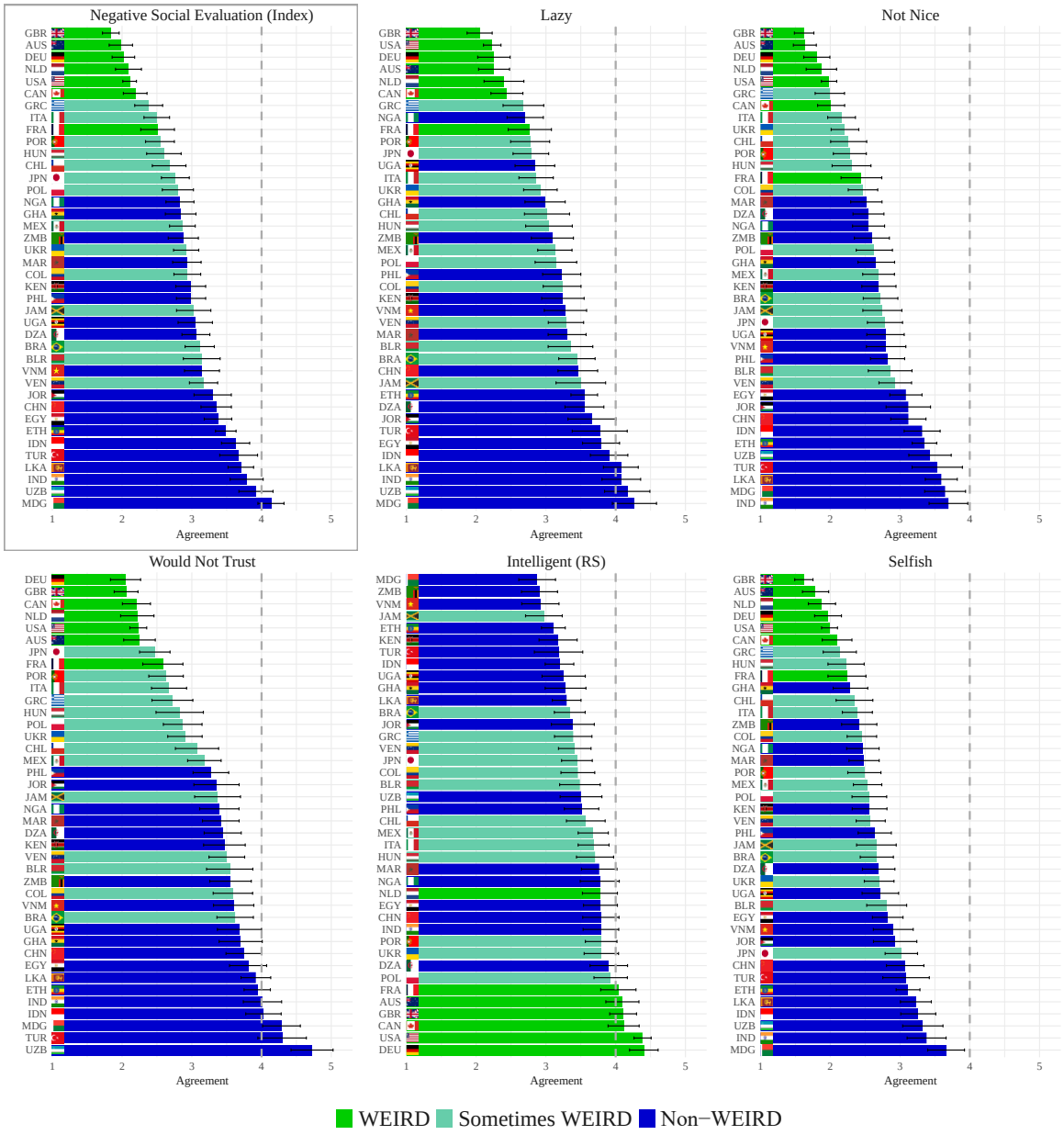
Measures of effort in Study 5a (N = 7,823 independent participants recruited on seven crowd-sourcing platforms). The top panel presents the marginal means adjusted for laptop use for the proportion of participants who complete the whole data entry task; the middle panel shows the marginal means for total number of cells inputted (out of 132); the bottom panel shows the marginal means for the total number of cells inputted correctly (similarly out of 132). The mixed-effects logistic regression analysis showed a significant negative linear effect of the "WEIRD" country classification on the probability of completing the whole task, b = -1.035, z = -6.070, P < 0.001, 95% CI -1.370 to -0.701, indicating that participants from more "WEIRD" countries were less likely to complete the entire transcription. The quadratic term for the "WEIRD" classification was also significant and negative, b = -0.315, z = -2.494, P = 0.013, 95% CI -0.562 to -0.067, suggesting that the decline in completion rates among "WEIRD" countries and "Partly (Sometimes) WEIRD" countries was not linear across levels of the "WEIRD" classification. The mixed-effects logistic regression on total cells inputted also showed a significant negative linear effect of the "WEIRD" classification, b = -0.374, z = -4.904, P < 0.001, 95% CI -0.524 to -0.225, indicating that countries that are more "WEIRD" completed fewer cells. The quadratic term for the "WEIRD" classification also had a significant negative effect, b = -0.132, z = -2.520, P = 0.016, 95% CI -0.234 to -0.029, suggesting that the decline in the number of cells inputted was not linear across "WEIRD" and "Partly (Sometimes) WEIRD" countries. The linear effect for cells inputted correctly was similarly negative and significant, b = -0.385, z = -5.116, P < 0.001, 95% CI -0.533 to -0.238; the same applies to the quadratic term for the "WEIRD" classification, b = -0.135, z = -2.628, P = 0.012, 95% CI -0.236 to -0.034. Error bars are 95% CIs.

Additional Fig. 8: **Moral Evaluations of Anticipation of Reciprocity Upon Completing the Task in Study 5A.**



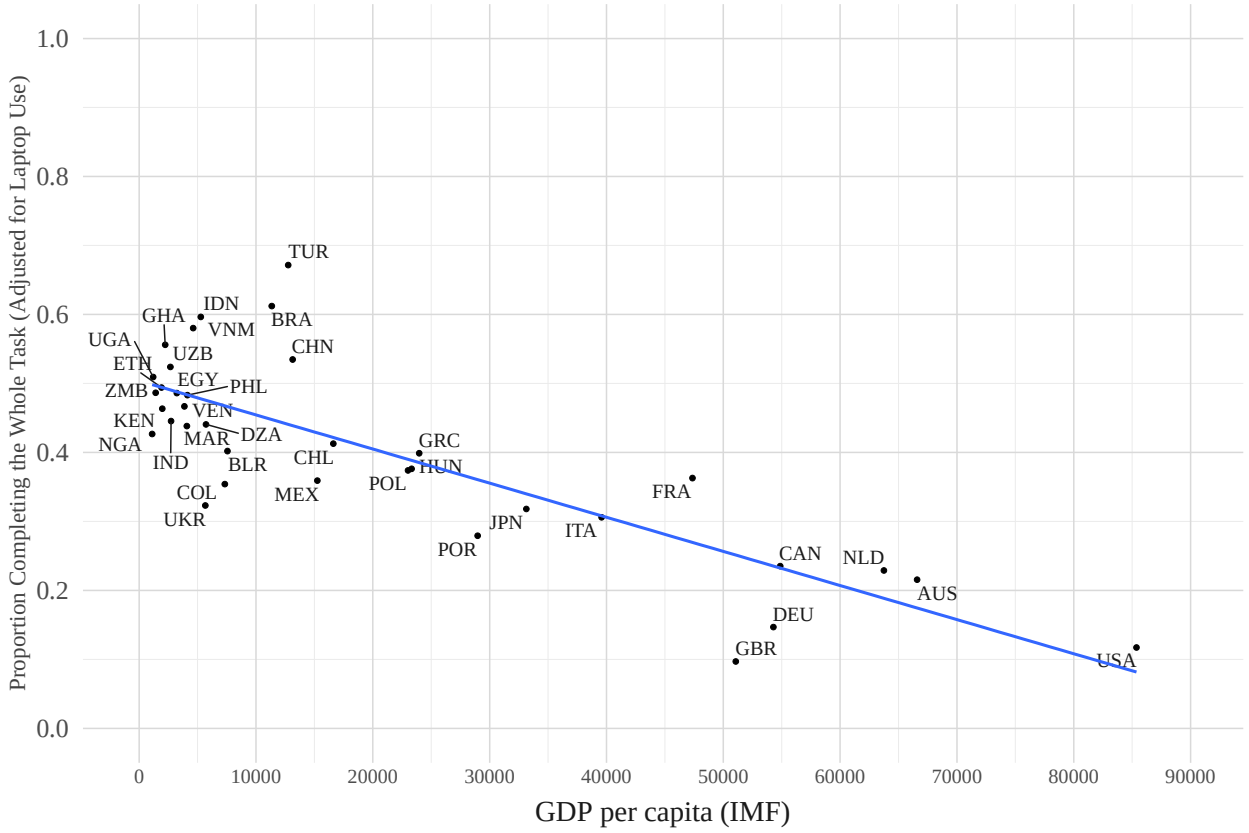
Moral evaluations of working beyond the contractual minimum required ("The morally right thing to do was to spend more than 5 minutes on the task") among participants Study 5a (N = 7,823 independent participants recruited on seven crowd-sourcing platforms). The linear term for "WEIRD" shows a significant negative effect, indicating that participants from "WEIRD" countries are less likely to view completing the task as moral, b = -0.211, t = -3.611, P = 0.001, 95% CI -0.326 to -0.097, indicating that individuals from "WEIRD" countries were less likely to agree that it is morally right to complete the task (Appendix Table S20). The quadratic term was not significant, b = -0.001, t = -0.028, P = 0.978, 95% CI -0.079 to 0.076, suggesting that the decrease in moral perceptions was not steeper in "WEIRD" or in "Partly (Sometimes) WEIRD."

Additional Fig. 9: **Social Evaluations of Those Who Do Not Complete the Whole Task in Study 5A.**



Negative social evaluations (index) of those who do not complete the task in Study 5a (N = 7,823 independent participants recruited on seven crowd-sourcing platforms). Controlling for whether participants themselves complete the task, the linear term for "WEIRD" shows a significant negative effect, indicating that participants from more "WEIRD" countries were less likely to have negative social evaluations of those who do not complete the task, b = -0.160, t = -2.258, P = 0.027, 95% CI -0.298 to -0.021 (Appendix Table S21). Conversely, the quadratic term for "WEIRD" was not significant, b = 0.008, t = 0.164, P = 0.870, 95% CI -0.091 to 0.108, suggesting that the negative social evaluations decreased linearly along the "WEIRD" classification.

Additional Fig. 10: **Study 5A, Correlation Between GDP per capita and Marginal Completion Rates (Adjusted for Device Type).**



Correlation between marginal completion rates of the data entry task in Study 5A, grouped by country, and GDP per capita of that country taken from the International Monetary Fund data (International Monetary Fund, 2023).

Additional Table 1: **Study 5A, Correlations Between Marginal Completion Rates and Country-Level Metrics.**

Variable	Correlation	CI Lower	CI Upper	P Value
GDP per capita (IMF)	-0.81	-0.9	-0.67	< 0.001
Collectivism (Pelham)	0.76	0.58	0.86	< 0.001
Responsibilism (Talhelm)	0.76	0.59	0.87	< 0.001
Absence of Corruption (WJP/Sub-Scale of Rule of Law)	-0.73	-0.85	-0.54	< 0.001
Rule of Law Index (WJP)	-0.72	-0.84	-0.53	< 0.001
Cultural Distance from the US (Muthukrishna)	0.71	0.5	0.84	< 0.001
Individualism (Hofstede)	-0.71	-0.84	-0.52	< 0.001
Regulatory Enforcement (WJP/Sub-Scale of Rule of Law)	-0.71	-0.84	-0.52	< 0.001
Fragile State Index (FFP)	0.65	0.43	0.8	< 0.001
Personal Freedom (FI)	-0.65	-0.8	-0.42	< 0.001
Human Capital Index (WB)	-0.65	-0.8	-0.43	< 0.001
Share Informal Economy (ILO)	0.61	0.33	0.79	< 0.001
Economic Freedom (FI)	-0.55	-0.74	-0.29	< 0.001
Power Distance (Hofstede)	0.49	0.21	0.7	0.002
Tightness (Jackson)	0.46	0.11	0.71	0.015
Inequality, Gini (WB)	0.41	0.11	0.63	0.009
Uncertainty Avoidance (Hofstede)	-0.06	-0.37	0.26	0.705

Correlation table between marginal completion rates of the data entry task in Study 5A, grouped by country, and several country-level metrics.

APPENDIX

REGRESSION TABLES FOR STUDIES 1–4

Appendix Table S1: Study 1, Pooled Monetary Versus Pooled Non-Monetary Conditions in India and the US.

	b	*SE*	*t*	*P*	95% CI
Monetary Incentive	159.85	54.84	2.91	0.004	[52.35, 267.35]
US	195.69	36.04	5.43	< 0.001	[125.05, 266.34]
Female	-143.21	18.82	-7.61	< 0.001	[-180.10, -106.33]
Education	-101.11	18.95	-5.33	< 0.001	[-138.26, -63.95]
Age	-69.92	7.36	-9.50	< 0.001	[-84.35, -55.49]
Monetary Incentive × US	170.56	58.48	2.92	0.004	[55.91, 285.21]

Regression output for the multiple linear regression model on the number of button-presses, reported in the Results section of Study 1. Monetary Incentive is a binary variable for whether the participants were in one of the monetary (pay-for-effort) conditions. US is a binary variable for whether the participants were part of the US sample. Female is a categorical variable for gender. Age is measured on a scale, where 1 = "18–24"; 2 = "25–30"; 3 = "31–40"; 4 = "41–50"; 5 = "51–64"; 6 = "65 or higher." Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment.

Appendix Table S2: **Study 2a, Monetary Versus Pooled Non-Monetary Conditions in the UK and China.**

	b	*SE*	*t*	*P*	95% CI
Monetary Incentive	8.79	2.83	3.11	0.002	[3.25, 14.33]
UK	2.18	2.76	0.79	0.429	[-3.22, 7.58]
Female	1.13	1.89	0.6	0.551	[-2.58, 4.84]
Non-Binary	10.96	12.66	0.87	0.387	[-13.88, 35.79]
Education	-5.03	2.23	-2.25	0.024	[-9.41, -0.65]
Age	-2.73	1.23	-2.22	0.027	[-5.14, -0.32]
Monetary Incentive × UK	38.40	3.98	9.65	< 0.001	[30.59, 46.20]

Regression output for the multiple linear regression model on the number of images rated in the Study 2a samples from China *(N* = 1,086 recruited on social media) and the UK (*N* = 1,067 recruited on Prolific). Monetary Incentive is a binary variable for whether the participants were assigned to the monetary condition. UK is a binary variable for whether the participants were part of the UK sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Age is a continuous variable and is scaled.

***Appendix Table S3*: Study 2b, Monetary Versus Pooled Non-Monetary Conditions in the US and Mexico.**

	b	*SE*	*t*	*P*	95% CI
Mexico and the US (Same Nominal Pay)					
Monetary Incentive	27.49	2.92	9.42	< 0.001	[21.76, 33.21]
US	-37.51	2.62	-14.32	< 0.001	[-42.64, -32.37]
Female	-4.95	1.96	-2.52	0.012	[-8.80, -1.09]
Non-Binary	11.68	6.46	1.81	0.071	[-1.00, 24.36]
Education	-5.91	2.14	-2.76	0.006	[-10.12, -1.71]
Age	-1.36	1.13	-1.21	0.227	[-3.57, 0.85]
Monetary Incentive × US	13.93	4.07	3.42	< 0.001	[5.94, 21.92]
Mexico and the US (Same Subjective Pay)					
Monetary Incentive	27.49	2.95	9.31	< 0.001	[21.70, 33.28]
US	-36.21	2.61	-13.87	< 0.001	[-41.33, -31.09]
Female	-5.57	1.99	-2.80	< 0.001	[-9.47, -1.66]
Non-Binary	11.12	6.49	1.71	0.087	[-1.61, 23.85]
Education	-5.39	2.16	-2.50	0.013	[-9.61, -1.16]
Age	-2.76	1.12	-2.47	0.014	[-4.95, -0.56]
Monetary Incentive × US	19.28	4.09	4.71	< 0.001	[11.26, 27.30]

Regression output for the multiple linear regression on the number of images rated in the Study 2b sample from Mexico (N = 1,053 recruited on Prolific) and each of the two Study 2b samples in the US: one with the same nominal pay (N = 1,098 on Prolific) as in Mexico and the other with the same subjective (Becker et al., 1964) pay (N = 1,122 recruited on Prolific) as in Mexico. Monetary Incentive is a binary variable for whether the participants were assigned to the monetary condition. US is a binary variable for whether the participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Age is a continuous variable and is scaled.

Appendix Table S4: **Study 2c, Monetary Versus Pooled Non-Monetary Conditions in the US and South Africa.**

	b	*SE*	*t*	*P*	95% CI
South Africa and the US (Same Nominal Pay)					
Monetary Incentive	25.29	2.88	8.78	< 0.001	[19.64, 30.93]
US	-8.97	3.02	-2.97	0.003	[-14.90, -3.04]
Female	0.49	2.07	0.24	0.813	[-3.57, 4.55]
Non-Binary	-7.13	7.37	-0.97	0.333	[-21.59, 7.32]
Education	-6.98	2.08	-3.36	0.001	[-11.05, -2.90]
Age	-2.97	1.11	-2.68	0.007	[-5.14, -0.80]
Monetary Incentive × US	7.98	4.05	1.97	0.049	[0.03, 15.92]
South Africa and the US (Same Subjective Pay)					
Monetary Incentive	25.22	2.92	8.63	< 0.001	[19.48, 30.95]
US	-10.78	3.02	-3.57	< 0.001	[-16.70, -4.86]
Female	-2.09	2.09	-1.00	0.318	[-6.19, 2.01]
Non-Binary	-13.06	7.92	-1.65	0.099	[-28.59, 2.47]
Education	-4.21	2.11	-1.99	0.046	[-8.36, -0.07]
Age	-1.02	1.11	-0.92	0.359	[-3.20, 1.16]
Monetary Incentive × US	15.40	4.12	3.74	< 0.001	[7.33, 23.48]

Regression output for the multiple linear regression on the number of images rated in the Study 2c sample from South Africa (N = 649 recruited on Prolific) and each of the two Study 2c samples in the US: one with the same nominal pay (N = 662 recruited on Prolific) as in South Africa and the other with the same subjective (Becker et al., 1964) pay (N = 662 recruited on Prolific) as in South Africa. Monetary Incentive is a binary variable for whether the participants were assigned to the monetary condition. US is a binary variable for whether the participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Age is a continuous variable and is scaled.

Appendix Table S5: **Study 3a, Norm Versus Minimal Pay Condition in India and the US.**

	b	*SE*	*t*	*P*	95% CI
Monetary Incentive	0.83	4.28	0.19	0.847	[-7.57, 9.22]
US	-26.03	4.56	-5.71	< 0.001	[-34.99, -17.08]
Female	2.59	3.23	0.80	0.422	[-3.75, 8.94]
Non-Binary	21.16	11.10	1.91	0.057	[-0.62, 42.95]
Education	-4.45	3.87	-1.15	0.250	[-12.05, 3.14]
Age	-0.42	1.50	-0.28	0.781	[-3.36, 2.52]
Monetary Incentive × US	13.77	5.95	2.31	0.021	[2.09, 25.45]

Regression output for the multiple linear regression on the number of images rated in the Study 3a samples from India (N = 352 recruited on MTurk) and the US (N = 382 recruited on Prolific). Monetary Incentive is a binary variable for whether the participants were assigned to the monetary (minimal pay) condition. US is a binary variable for whether the participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Age is a continuous variable and is scaled.

Appendix Table S6: **Study 4, Norm Versus Monetary Condition in India, by Language Prime (English or Hindi).**

	b	*SE*	*t*	*P*	95% CI
Monetary Incentive	12.63	2.27	5.56	< 0.001	[8.17, 17.09]
English	-1.74	2.28	-0.76	0.445	[-6.22, 2.74]
Monetary Incentive × English	10.69	3.23	3.31	0.001	[4.35, 17.02]

Regression output for the multiple linear regression on the number of images rated in the sample of bilingual participants from India (N = 2,065 recruited on Facebook), reported in the Results section of Study 4. Monetary Incentive is a binary variable for whether participants were assigned to the monetary condition. English is a binary variable for whether participants were assigned to complete the study in English (as compared to in Hindi).

Appendix Table S7: **Study 2a, Monetary Versus Individual Non-Monetary Conditions in the UK and China.**

	b	*SE*	*t*	*P*	95% CI
Flat Fee	-12.23	3.25	-3.76	< 0.001	[-18.60, -5.86]
Norm	-5.63	3.19	-1.77	0.078	[-11.87, 0.62]
UK	40.55	3.56	11.40	< 0.001	[33.58, 47.53]
Female	1.26	1.87	0.67	0.500	[-2.41, 4.94]
Non-Binary	12.66	12.55	1.01	0.313	[-11.95, 37.27]
Education	-5.07	2.21	-2.29	0.022	[-9.41, -0.74]
Age	-2.72	1.22	-2.23	0.026	[-5.11, -0.33]
Flat Fee × UK	-44.88	4.57	-9.82	< 0.001	[-53.84, -35.91]
Norm x UK	-31.55	4.53	-6.96	< 0.001	[-40.44, -22.67]

Regression output for the multiple linear regression on the number of images rated in individual conditions in the Study 2a samples from China (N = 1,086 recruited on social media) and the UK (N = 1,067 recruited on Prolific). Flat Fee is a binary variable for whether participants were assigned to the flat fee condition. Norm is a binary variable for whether participants were assigned to the norm condition. UK is a binary variable for whether participants were part of the UK sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who have either completed or are currently pursuing an undergraduate degree or higher. Age is a continuous variable and is scaled.

Appendix Table S8: **Study 2b, Monetary Versus Individual Non-Monetary Conditions in the US and Mexico.**

	b	*SE*	*t*	*P*	95% CI
Mexico and the US (Same Nominal Pay)					
Flat Fee	-38.24	3.30	-11.59	< 0.001	[-44.71, -31.77]
Norm	-16.42	3.32	-4.94	< 0.001	[-22.94, -9.90]
US	-23.46	3.51	-6.69	< 0.001	[-30.34, -16.58]
Female	-4.60	1.94	-2.37	0.018	[-8.40, -0.80]
Non-Binary	11.65	6.37	1.83	0.068	[-0.84, 24.15]
Education	-5.87	2.11	-2.78	0.006	[-10.02, -1.73]
Age	-1.51	1.11	-1.36	0.174	[-3.69, 0.67]
Flat Fee × US	-10.32	4.61	-2.24	0.025	[-19.37, -1.27]
Norm x US	-17.69	4.64	-3.81	< 0.001	[-26.80, -8.58]
Mexico and the US (Same Subjective Pay)					
Flat Fee	-38.22	3.35	-11.42	< 0.001	[-44.79, -31.65]
Norm	-16.39	3.37	-4.86	< 0.001	[-23.01, -9.78]
US	-16.74	3.51	-4.77	< 0.001	[-23.62, -9.86]
Female	-5.56	1.97	-2.82	0.005	[-9.43, -1.70]
Non-Binary	11.28	6.42	1.76	0.079	[-1.31, 23.87]
Education	-4.87	2.13	-2.28	0.023	[-9.05, -0.69]
Age	-2.88	1.11	-2.60	0.009	[-5.05, -0.71]
Flat Fee × US	-13.41	4.66	-2.88	0.004	[-22.55, -4.27]
Norm x US	-25.55	4.68	-5.47	< 0.001	[-34.72, -16.38]

Regression output for the multiple linear regression on the number of images rated in individual conditions in the Study 2b sample from Mexico *(N* = 1,053 recruited on Prolific) and each of the two Study 2b samples in the US: one with the same nominal pay (*N* = 1,098 on Prolific) as in Mexico and one with the same subjective (Becker et al., 1964) pay (*N* = 1,122 recruited on Prolific) as in Mexico. Flat Fee is a binary variable for whether participants were assigned to the flat fee condition. Norm is a binary variable for whether participants were assigned to the norm condition. US is a binary variable for whether participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who have either completed or are currently pursuing an undergraduate degree or higher. Age is a continuous variable and is scaled.

REGRESSION TABLES FOR STUDIES 5A AND 5B

Appendix Table S9: **Study 5a, Explicitly No Pay, Completion Rates Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-1.035	0.171	-6.070	< 0.001	[-1.370, -0.701]
WEIRD, Quadratic	-0.315	0.126	-2.494	0.013	[-0.562, -0.067]
Male	0.124	0.056	2.230	0.026	[0.015, 0.234]
Age	0.040	0.033	1.197	0.231	[-0.025, 0.105]
Employed	0.031	0.060	0.523	0.601	[-0.086, 0.148]
Student	-0.021	0.069	-0.298	0.766	[-0.156, 0.115]
Education	-0.070	0.029	-2.447	0.014	[-0.127, -0.014]
Urban	-0.078	0.060	-1.306	0.192	[-0.196, 0.039]
Majority Race	-0.290	0.071	-4.075	< 0.001	[-0.430, -0.151]
SES	0.019	0.029	0.669	0.503	[-0.037, 0.076]
Income	0.042	0.039	1.077	0.282	[-0.035, 0.119]
Platform: MTurk	0.234	0.210	1.114	0.265	[-0.178, 0.646]
Platform: Toloka	0.328	0.160	2.049	0.040	[0.014, 0.642]
Platform: BeSample	0.650	0.173	3.765	< 0.001	[0.312, 0.988]
Platform: Aishiyan	0.621	0.380	1.636	0.102	[-0.123, 1.366]
Platform: Lancers	-0.206	0.412	-0.500	0.617	[-1.014, 0.602]
Phone	-1.002	0.087	-11.472	< 0.001	[-1.173, -0.831]

The regression output for the mixed-effects logistic regression model on task completion in Study 5a. The dependent variable is task completion, which is a binary and indicates whether participants completed the entire data entry sheet (1) or not (0). The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a

medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S10: **Study 5a, Explicitly No Pay, Total Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.374	0.076	-4.904	< 0.001	[-0.524, -0.225]
WEIRD, Quadratic	-0.132	0.052	-2.520	0.016	[-0.234, -0.029]
Male	0.027	0.024	1.110	0.267	[-0.021, 0.074]
Age	-0.020	0.014	-1.423	0.155	[-0.048, 0.008]
Employed	0.013	0.026	0.494	0.621	[-0.038, 0.064]
Student	-0.008	0.030	-0.261	0.794	[-0.067, 0.051]
Education	-0.013	0.012	-1.030	0.303	[-0.037, 0.012]
Urban	-0.021	0.026	-0.799	0.424	[-0.072, 0.030]
Majority Race	-0.124	0.031	-3.993	< 0.001	[-0.185, -0.063]
SES	0.004	0.013	0.314	0.753	[-0.021, 0.029]
Income	0.035	0.016	2.133	0.033	[0.003, 0.067]
Platform: MTurk	0.125	0.088	1.422	0.155	[-0.047, 0.297]
Platform: Toloka	0.020	0.074	0.270	0.787	[-0.124, 0.164]
Platform: BeSample	0.179	0.079	2.277	0.023	[0.025, 0.333]
Platform: Aishiyan	0.306	0.194	1.579	0.123	[-0.074, 0.686]
Platform: Lancers	-0.042	0.209	-0.202	0.841	[-0.451, 0.367]
Phone	-0.671	0.039	-17.129	< 0.001	[-0.748, -0.594]

The regression output for the mixed-effects regression model on total number of cells inputted (out of 132) in Study 5a. The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes

for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S11: **Study 5a, Explicitly No Pay, Correct Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.385	0.075	-5.116	< 0.001	[-0.533, -0.238]
WEIRD, Quadratic	-0.135	0.051	-2.628	0.012	[-0.236, -0.034]
Male	0.029	0.024	1.215	0.224	[-0.018, 0.076]
Age	-0.014	0.014	-0.976	0.329	[-0.042, 0.014]
Employed	-0.001	0.026	-0.047	0.962	[-0.052, 0.049]
Student	-0.024	0.030	-0.790	0.430	[-0.083, 0.035]
Education	-0.010	0.012	-0.805	0.421	[-0.034, 0.014]
Urban	-0.007	0.026	-0.276	0.783	[-0.058, 0.044]
Majority Race	-0.103	0.031	-3.335	0.001	[-0.164, -0.043]
SES	-0.001	0.013	-0.079	0.937	[-0.026, 0.024]
Income	0.040	0.016	2.445	0.015	[0.008, 0.072]
Platform: MTurk	0.072	0.088	0.826	0.409	[-0.099, 0.244]
Platform: Toloka	0.026	0.073	0.357	0.721	[-0.117, 0.169]
Platform: BeSample	0.154	0.078	1.975	0.049	[0.001, 0.307]
Platform: Aishiyan	0.194	0.188	1.032	0.309	[-0.175, 0.564]
Platform: Lancers	-0.096	0.208	-0.461	0.647	[-0.503, 0.311]
Phone	-0.658	0.038	-17.195	< 0.001	[-0.733, -0.583]

The regression output for the mixed-effects regression model on total number of cells inputted correctly (out of 132) in Study 5a. The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model

includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S12: **Study 5a, Explicitly No Pay, Completion Rates Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries for Workers on Laptops.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.992	0.178	-5.559	< 0.001	[-1.342, -0.642]
WEIRD, Quadratic	-0.111	0.116	-0.957	0.338	[-0.339, 0.117]
Male	0.173	0.070	2.475	0.013	[0.036, 0.309]
Age	0.060	0.041	1.450	0.147	[-0.021, 0.140]
Employed	0.133	0.075	1.782	0.075	[-0.013, 0.279]
Student	-0.055	0.089	-0.612	0.541	[-0.229, 0.120]
Education	-0.049	0.036	-1.370	0.171	[-0.118, 0.021]
Urban	-0.107	0.076	-1.416	0.157	[-0.256, 0.041]
Majority Race	-0.310	0.092	-3.379	0.001	[-0.490, -0.130]
SES	0.014	0.037	0.390	0.696	[-0.058, 0.086]
Income	0.061	0.049	1.244	0.214	[-0.035, 0.156]
Platform: MTurk	0.093	0.220	0.422	0.673	[-0.339, 0.525]
Platform: Toloka	0.090	0.178	0.505	0.613	[-0.259, 0.439]
Platform: BeSample	0.393	0.204	1.930	0.054	[-0.006, 0.792]
Platform: Aishiyan	-0.022	0.542	-0.041	0.967	[-1.085, 1.041]
Platform: Lancers	-0.356	0.418	-0.851	0.395	[-1.175, 0.463]

The regression output for the mixed-effects logistic regression model on task completion in Study 5a for people completing the task on laptops (N = 4,923 independent participants recruited on seven different crowd-sourcing platforms). The dependent variable is task completion, which is a binary and indicates whether participants completed the entire data entry sheet (1) or not (0). The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. The model includes random intercepts for country. All non-binary variables are scaled.

Appendix Table S13: **Study 5a, Explicitly No Pay, Completion Rates Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries for Workers on Phones.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-1.363	0.279	-4.889	< 0.001	[-1.910, -0.817]
WEIRD, Quadratic	-0.717	0.161	-4.458	< 0.001	[-1.032, -0.402]
Male	0.037	0.092	0.397	0.692	[-0.144, 0.217]
Age	0.018	0.056	0.321	0.748	[-0.091, 0.127]
Employed	-0.136	0.099	-1.377	0.169	[-0.330, 0.058]
Student	0.020	0.109	0.181	0.857	[-0.195, 0.234]
Education	-0.115	0.048	-2.391	0.017	[-0.209, -0.021]
Urban	-0.026	0.098	-0.263	0.793	[-0.218, 0.166]
Majority Race	-0.257	0.109	-2.354	0.019	[-0.471, -0.043]
SES	0.034	0.046	0.723	0.470	[-0.057, 0.124]
Income	-0.028	0.070	-0.403	0.687	[-0.167, 0.110]
Platform: MTurk	-0.122	0.816	-0.149	0.881	[-1.721, 1.477]
Platform: Toloka	0.681	0.250	2.723	0.006	[0.191, 1.170]
Platform: BeSample	1.032	0.262	3.932	< 0.001	[0.518, 1.547]
Platform: Aishiyan	1.146	0.376	3.048	0.002	[0.409, 1.882]
Platform: Lancers	0.523	0.755	0.693	0.489	[-0.957, 2.003]

The regression output for the mixed-effects logistic regression model on task completion in Study 5a for people completing the task on their phones (N = 2,900 independent participants recruited on seven different crowd-sourcing platforms). The dependent variable is task completion, which is a binary and indicates whether participants completed the entire data entry sheet (1) or not (0). The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. The model includes random intercepts for country. All non-binary variables are scaled.

Appendix Table S14: **Study 5b, Ambiguous Pay, Completion Rates Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.852	0.146	-5.852	< 0.001	[-1.137, -0.567]
WEIRD, Quadratic	-0.023	0.100	-0.233	0.815	[-0.220, 0.173]
Male	0.085	0.055	1.528	0.127	[-0.024, 0.193]
Age	-0.075	0.033	-2.279	0.023	[-0.139, -0.010]
Employed	0.003	0.059	0.056	0.955	[-0.113, 0.120]
Student	0.006	0.069	0.081	0.935	[-0.129, 0.140]
Education	-0.017	0.029	-0.604	0.546	[-0.074, 0.039]
Urban	-0.103	0.060	-1.712	0.087	[-0.221, 0.015]
Majority Race	-0.237	0.072	-3.293	0.001	[-0.379, -0.096]
SES	0.031	0.029	1.062	0.288	[-0.026, 0.087]
Income	0.003	0.037	0.079	0.937	[-0.070, 0.076]
Platform: MTurk	-0.014	0.183	-0.076	0.939	[-0.372, 0.344]
Platform: Toloka	0.096	0.146	0.656	0.512	[-0.191, 0.383]
Platform: BeSample	0.528	0.163	3.248	0.001	[0.209, 0.847]
Platform: Aishiyan	0.591	0.402	1.472	0.141	[-0.196, 1.379]
Platform: Lancers	0.171	0.354	0.485	0.628	[-0.522, 0.865]
Phone	-0.999	0.076	-13.175	< 0.001	[-1.148, -0.851]

The regression output for the mixed-effects logistic regression model on task completion in Study 5b. The dependent variable is task completion, which is a binary and indicates whether participants completed the entire data entry sheet (1) or not (0). The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model

includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S15: **Study 5b, Ambiguous Pay, Total Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.342	0.069	-4.983	< 0.001	[-0.477, -0.208]
WEIRD, Quadratic	-0.003	0.047	-0.075	0.941	[-0.095, 0.088]
Male	0.023	0.024	0.974	0.330	[-0.024, 0.071]
Age	-0.050	0.014	-3.566	< 0.001	[-0.078, -0.023]
Employed	0.009	0.026	0.346	0.729	[-0.041, 0.059]
Student	-0.006	0.030	-0.204	0.838	[-0.064, 0.052]
Education	-0.003	0.012	-0.256	0.798	[-0.028, 0.021]
Urban	-0.032	0.026	-1.245	0.213	[-0.083, 0.019]
Majority Race	-0.119	0.031	-3.859	< 0.001	[-0.180, -0.059]
SES	0.020	0.012	1.605	0.108	[-0.004, 0.044]
Income	0.010	0.016	0.646	0.518	[-0.021, 0.042]
Platform: MTurk	-0.034	0.080	-0.421	0.674	[-0.190, 0.123]
Platform: Toloka	-0.091	0.067	-1.345	0.180	[-0.223, 0.041]
Platform: BeSample	0.114	0.073	1.568	0.118	[-0.028, 0.256]
Platform: Aishiyan	0.242	0.193	1.253	0.216	[-0.137, 0.621]
Platform: Lancers	0.125	0.176	0.707	0.484	[-0.221, 0.471]
Phone	-0.655	0.040	-16.294	< 0.001	[-0.734, -0.576]

The regression output for the mixed-effects regression model on total number of cells inputted (out of 132) in Study 5b. The predictors include "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes

for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S16: **Study 5b, Ambiguous Pay, Correct Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.311	0.072	-4.302	< 0.001	[-0.452, -0.169]
WEIRD, Quadratic	-0.004	0.050	-0.076	0.940	[-0.102, 0.094]
Male	0.021	0.024	0.868	0.385	[-0.026, 0.069]
Age	-0.050	0.014	-3.541	< 0.001	[-0.078, -0.022]
Employed	-0.005	0.026	-0.178	0.858	[-0.055, 0.046]
Student	-0.017	0.030	-0.574	0.566	[-0.076, 0.042]
Education	0.006	0.013	0.455	0.649	[-0.019, 0.030]
Urban	-0.033	0.026	-1.273	0.203	[-0.085, 0.018]
Majority Race	-0.099	0.031	-3.177	0.001	[-0.161, -0.038]
SES	0.014	0.013	1.135	0.256	[-0.010, 0.039]
Income	0.013	0.016	0.819	0.413	[-0.019, 0.045]
Platform: MTurk	-0.033	0.081	-0.409	0.683	[-0.191, 0.125]
Platform: Toloka	-0.026	0.070	-0.375	0.708	[-0.162, 0.110]
Platform: BeSample	0.138	0.074	1.860	0.064	[-0.007, 0.284]
Platform: Aishiyan	0.260	0.201	1.290	0.203	[-0.135, 0.654]
Platform: Lancers	0.124	0.188	0.661	0.513	[-0.244, 0.493]
Phone	-0.657	0.038	-17.501	< 0.001	[-0.73, -0.583]

The regression output for the mixed-effects regression model on total number of cells inputted correctly (out of 132) in Study 5a. The predictors include "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model

includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S17: **Studies 5a and 5b, Interaction Between Pay Conditions, Completion Rates Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
Condition, Ambiguous	0.436	0.041	10.565	< 0.001	[0.355, 0.517]
WEIRD, Linear	-0.996	0.147	-6.777	< 0.001	[-1.284, -0.708]
WEIRD, Quadratic	-0.239	0.116	-2.061	0.039	[-0.466, -0.012]
Male	0.105	0.039	2.659	0.008	[0.028, 0.182]
Age	-0.014	0.023	-0.595	0.552	[-0.060, 0.032]
Employed	0.011	0.042	0.255	0.799	[-0.072, 0.093]
Student	-0.010	0.049	-0.214	0.830	[-0.106, 0.085]
Education	-0.042	0.020	-2.071	0.038	[-0.082, -0.002]
Urban	-0.098	0.043	-2.291	0.022	[-0.181, -0.014]
Majority Race	-0.258	0.051	-5.025	< 0.001	[-0.358, -0.157]
SES	0.022	0.020	1.084	0.278	[-0.018, 0.062]
Income	0.023	0.027	0.854	0.393	[-0.030, 0.077]
Platform: MTurk	0.107	0.141	0.758	0.448	[-0.170, 0.384]
Platform: Toloka	0.204	0.122	1.669	0.095	[-0.036, 0.444]
Platform: BeSample	0.568	0.130	4.372	< 0.001	[0.313, 0.822]
Platform: Aishiyan	0.525	0.372	1.414	0.157	[-0.203, 1.254]
Platform: Lancers	-0.009	0.357	-0.025	0.980	[-0.709, 0.691]
Phone	-0.997	0.069	-14.500	0.000	[-1.132, -0.862]
Cond, Ambiguous x WEIRD, Linear	0.133	0.075	1.784	0.074	[-0.013, 0.280]
Cond, Ambiguous x WEIRD, Quadratic	0.168	0.068	2.467	0.014	[0.035, 0.302]

The regression output for the mixed-effects logistic regression model on task completion in Studies 5a and 5b. The dependent variable is task completion, which is a binary and indicates whether participants completed the entire data entry sheet (1) or not (0). The predictors include Condition (Explicitly No Pay or Ambiguous Pay); "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to

$59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S18: **Studies 5a and 5b, Interaction Between Pay Conditions, Total Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
Condition, Ambiguous	0.178	0.017	10.386	< 0.001	[0.145, 0.212]
WEIRD, Linear	-0.351	0.066	-5.312	< 0.001	[-0.481, -0.222]
WEIRD, Quadratic	-0.105	0.050	-2.113	0.041	[-0.203, -0.008]
Male	0.024	0.017	1.435	0.151	[-0.009, 0.058]
Age	-0.034	0.010	-3.393	0.001	[-0.054, -0.014]
Employed	0.009	0.018	0.477	0.633	[-0.027, 0.045]
Student	-0.009	0.021	-0.404	0.686	[-0.050, 0.033]
Education	-0.008	0.009	-0.859	0.390	[-0.025, 0.010]
Urban	-0.028	0.018	-1.526	0.127	[-0.064, 0.008]
Majority Race	-0.120	0.022	-5.404	< 0.001	[-0.163, -0.076]
SES	0.011	0.009	1.251	0.211	[-0.006, 0.029]
Income	0.023	0.012	2.013	0.044	[0.001, 0.046]
Platform: MTurk	0.052	0.060	0.868	0.385	[-0.066, 0.170]
Platform: Toloka	-0.019	0.056	-0.346	0.730	[-0.128, 0.090]
Platform: BeSample	0.157	0.058	2.698	0.007	[0.043, 0.271]
Platform: Aishiyan	0.268	0.180	1.488	0.144	[-0.085, 0.622]
Platform: Lancers	0.051	0.182	0.283	0.779	[-0.304, 0.407]
Phone	-0.660	0.034	-19.250	< 0.001	[-0.727, -0.593]
Cond, Ambiguous x WEIRD, Linear	0.017	0.030	0.551	0.582	[-0.043, 0.076]
Cond, Ambiguous x WEIRD, Quadratic	0.081	0.029	2.749	0.006	[0.023, 0.138]

The regression output for the mixed-effects regression model on total number of cells inputted (out of 132) in Studies 5a and 5b. The predictors include Condition (Explicitly No Pay or Ambiguous Pay); "WEIRD, Linear" and WEIRD, Quadratic, which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous

("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S19: **Studies 5a and 5b, Interaction Between Pay Conditions, Correct Input Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
Condition, Ambiguous	0.171	0.017	9.966	< 0.001	[0.138, 0.205]
WEIRD, Linear	-0.347	0.067	-5.150	< 0.001	[-0.478, -0.215]
WEIRD, Quadratic	-0.111	0.051	-2.188	0.034	[-0.211, -0.012]
Male	0.024	0.017	1.430	0.153	[-0.009, 0.058]
Age	-0.031	0.010	-3.057	0.002	[-0.050, -0.011]
Employed	-0.005	0.018	-0.291	0.771	[-0.041, 0.031]
Student	-0.022	0.021	-1.030	0.303	[-0.063, 0.02]
Education	-0.002	0.009	-0.196	0.844	[-0.019, 0.016]
Urban	-0.022	0.018	-1.204	0.229	[-0.058, 0.014]
Majority Race	-0.099	0.022	-4.474	< 0.001	[-0.143, -0.056]
SES	0.005	0.009	0.611	0.541	[-0.012, 0.023]
Income	0.027	0.012	2.345	0.019	[0.004, 0.050]
Platform: MTurk	0.032	0.060	0.522	0.601	[-0.087, 0.150]
Platform: Toloka	0.022	0.056	0.391	0.696	[-0.088, 0.132]
Platform: BeSample	0.162	0.059	2.769	0.006	[0.047, 0.277]
Platform: Aishiyan	0.239	0.181	1.323	0.193	[-0.115, 0.593]
Platform: Lancers	0.013	0.189	0.071	0.944	[-0.357, 0.383]
Phone	-0.654	0.032	-20.160	< 0.001	[-0.718, -0.591]
Cond, Ambiguous x WEIRD, Linear	0.019	0.030	0.635	0.526	[-0.040, 0.078]
Cond, Ambiguous x WEIRD, Quadratic	0.080	0.029	2.717	0.007	[0.022, 0.137]

The regression output for the mixed-effects regression model on total number of cells inputted correctly (out of 132) in Studies 5a and 5b. The predictors include Condition (Explicitly No Pay or Ambiguous Pay); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a

continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes random intercepts and slopes for device type across countries, allowing the relationship between device type and task completion to vary by country. All non-binary variables are scaled.

Appendix Table S20: **Study 5a, Moral Evaluations of Completing the Task Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.211	0.058	-3.611	0.001	[-0.326, -0.097]
WEIRD, Quadratic	-0.001	0.040	-0.028	0.978	[-0.079, 0.076]
Completed the Whole Task	0.707	0.026	27.203	< 0.001	[0.657, 0.758]
Male	-0.001	0.022	-0.027	0.978	[-0.045, 0.043]
Age	-0.004	0.013	-0.293	0.769	[-0.030, 0.022]
Employed	0.009	0.024	0.392	0.695	[-0.038, 0.057]
Student	0.048	0.028	1.706	0.088	[-0.007, 0.103]
Education	-0.081	0.012	-7.059	< 0.001	[-0.104, -0.059]
Urban	-0.014	0.024	-0.598	0.550	[-0.062, 0.033]
Majority Race	-0.077	0.029	-2.698	0.007	[-0.133, -0.021]
SES	0.058	0.012	4.977	< 0.001	[0.035, 0.081]
Income	-0.054	0.015	-3.548	< 0.001	[-0.083, -0.024]
Platform: MTurk	0.163	0.078	2.083	0.037	[0.010, 0.316]
Platform: Toloka	0.253	0.058	4.338	< 0.001	[0.139, 0.367]
Platform: BeSample	0.287	0.064	4.452	< 0.001	[0.161, 0.413]
Platform: Aishiyan	0.395	0.141	2.805	0.007	[0.119, 0.670]
Platform: Lancers	0.004	0.138	0.032	0.974	[-0.265, 0.274]
Phone	0.137	0.024	5.694	< 0.001	[0.090, 0.184]
Completed the Whole Task x WEIRD, Linear	0.011	0.047	0.245	0.806	[-0.080, 0.103]
Completed the Whole Task x WEIRD, Quadratic	-0.080	0.042	-1.928	0.054	[-0.162, 0.001]

The regression output for the mixed-effects regression model on agreement with the item "The morally right thing to do was to spend more than 5 minutes on the task" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to

$39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled. Participants across countries who completed the whole task are significantly more likely to view it as the morally right thing to do, $b = 0.707$, $t = 27.203$, $P < 0.001$, 95% CI 0.657 to 0.758. The linear term for "WEIRD" shows a significant negative effect, indicating that participants from "WEIRD" countries are less likely to view completing the task as moral, $b = -0.211$, $t = -3.611$, $P = 0.001$, 95% CI -0.326 to -0.097, indicating that individuals from "WEIRD" countries were less likely to agree that it is morally right to complete the task. The quadratic term was not significant, $b = -0.001$, $t = -0.028$, $P = 0.978$, 95% CI -0.079 to 0.076, suggesting that a decrease in moral perceptions is not steeper for "WEIRD" compared to "Partly (Sometimes) WEIRD" above the linear term. The interaction between task completion and the linear term for "WEIRD" was not significant, indicating that among people who complete the whole task, the negative social evaluations were not significantly different for those from "WEIRD" countries, $b = 0.011$, $t = 0.245$, $P = 0.806$, 95% CI -0.080 to 0.103. Similarly, the interaction between task completion and the quadratic term for "WEIRD" was also not significant, suggesting a linear relationship along the levels of the "WEIRD" classification, $b = -0.080$, $t = -1.928$, $P = 0.054$, 95% CI -0.162 to 0.001.

Appendix Table S21: **Study 5a, Social Evaluations of People Who Do Not Complete the Whole Task Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.160	0.071	-2.258	0.027	[-0.298, -0.021]
WEIRD, Quadratic	0.008	0.051	0.164	0.870	[-0.091, 0.108]
Completed the Whole Task	0.828	0.026	31.871	< 0.001	[0.777, 0.879]
Male	0.084	0.023	3.617	< 0.001	[0.038, 0.129]
Age	-0.050	0.014	-3.670	< 0.001	[-0.077, -0.023]
Employed	-0.025	0.025	-0.972	0.331	[-0.074, 0.025]
Student	0.063	0.029	2.207	0.027	[0.007, 0.120]
Education	-0.051	0.012	-4.336	< 0.001	[-0.074, -0.028]
Urban	-0.065	0.025	-2.625	0.009	[-0.114, -0.017]
Majority Race	-0.089	0.030	-2.993	0.003	[-0.148, -0.031]
SES	0.052	0.012	4.321	< 0.001	[0.029, 0.076]
Income	-0.016	0.015	-1.086	0.278	[-0.046, 0.013]
Platform: MTurk	0.240	0.078	3.092	0.002	[0.088, 0.392]
Platform: Toloka	0.207	0.067	3.101	0.002	[0.076, 0.338]
Platform: BeSample	0.186	0.070	2.669	0.008	[0.049, 0.322]
Platform: Aishiyan	0.349	0.178	1.960	0.057	[-0.000, 0.699]
Platform: Lancers	0.266	0.174	1.530	0.135	[-0.075, 0.607]
Phone	0.071	0.025	2.859	0.004	[0.022, 0.119]
Completed the Whole Task x WEIRD, Linear	-0.169	0.046	-3.700	< 0.001	[-0.259, -0.080]
Completed the Whole Task x WEIRD, Quadratic	-0.152	0.043	-3.557	< 0.001	[-0.235, -0.068]

The regression output for the mixed-effects regression model on the index of negative social evaluations in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual

income (in USD) with the following answer choices: “Less than $5,000;” “$5,000 to $9,999;” "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled. Participants who completed the whole task are significantly more likely to have negative social evaluations of those who do not complete the task, $b = 0.828$, $t = 31.871$, $P < 0.001$, 95% CI 0.777 to 0.879. The linear term for “WEIRD” shows a significant negative effect, indicating that participants from “WEIRD” countries are less likely to have negative social evaluations of those who do not complete the task, $b = -0.160$, $t = -2.258$, $P = 0.027$, 95% CI -0.298 to -0.021. Conversely, the quadratic term for “WEIRD” was not significant, $b = 0.008$, $t = 0.164$, $P = 0.870$, 95% CI -0.091 to 0.108, suggesting that the most “WEIRD” countries do not decrease their negative social evaluations beyond the linear effect. The interaction between task completion and the linear term for “WEIRD” was significant, indicating that among people who complete the whole task, the negative social evaluations were less pronounced for those from “WEIRD” countries, $b = -0.169$, $t = -3.700$, $P < 0.001$, 95% CI -0.259 to -0.080. Similarly, the interaction between task completion and the quadratic term for “WEIRD” was significant, suggesting that this decline was not linear along the levels of the “WEIRD” classification, $b = -0.152$, $t = -3.557$, $P < 0.001$, 95% CI -0.235 to -0.068.

Appendix Table S22: **Study 5a, Expectations of More Pay for Completing the Whole Task Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.116	0.077	-1.507	0.136	[-0.266, 0.035]
WEIRD, Quadratic	-0.050	0.055	-0.903	0.371	[-0.157, 0.058]
Completed the Whole Task	0.339	0.027	12.584	< 0.001	[0.287, 0.392]
Male	0.013	0.023	0.561	0.575	[-0.033, 0.059]
Age	-0.035	0.014	-2.568	0.010	[-0.062, -0.008]
Employed	0.019	0.025	0.769	0.442	[-0.030, 0.068]
Student	0.104	0.029	3.582	< 0.001	[0.047, 0.161]
Education	-0.065	0.012	-5.425	< 0.001	[-0.089, -0.042]
Urban	-0.043	0.025	-1.695	0.090	[-0.092, 0.007]
Majority Race	-0.117	0.030	-3.873	< 0.001	[-0.176, -0.058]
SES	0.073	0.012	5.981	< 0.001	[0.049, 0.097]
Income	-0.053	0.016	-3.329	0.001	[-0.083, -0.022]
Platform: MTurk	0.470	0.083	5.637	< 0.001	[0.307, 0.634]
Platform: Toloka	0.369	0.071	5.224	< 0.001	[0.231, 0.507]
Platform: BeSample	0.453	0.075	6.018	< 0.001	[0.306, 0.601]
Platform: Aishiyan	-0.127	0.198	-0.640	0.526	[-0.515, 0.262]
Platform: Lancers	0.081	0.194	0.415	0.680	[-0.300, 0.461]
Phone	0.101	0.025	4.034	< 0.001	[0.052, 0.150]
Completed the Whole Task x WEIRD, Linear	-0.122	0.049	-2.508	0.012	[-0.217, -0.027]
Completed the Whole Task x WEIRD, Quadratic	0.012	0.043	0.272	0.786	[-0.073, 0.096]

The regression output for the mixed-effects regression model on agreement with the item "Working for longer than 5 minutes could have earned me more pay" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;"

"$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled.

Appendix Table S23: **Study 5a, Expectations of Future Work Opportunities for Completing the Whole Task Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.117	0.066	-1.787	0.078	[-0.246, 0.011]
WEIRD, Quadratic	0.034	0.045	0.752	0.456	[-0.054, 0.122]
Completed the Whole Task	0.297	0.027	10.793	< 0.001	[0.243, 0.350]
Male	0.046	0.024	1.922	0.055	[-0.001, 0.092]
Age	-0.045	0.014	-3.188	0.001	[-0.072, -0.017]
Employed	-0.020	0.026	-0.777	0.437	[-0.070, 0.030]
Student	0.065	0.030	2.195	0.028	[0.007, 0.123]
Education	-0.066	0.012	-5.435	< 0.001	[-0.090, -0.042]
Urban	-0.003	0.026	-0.117	0.907	[-0.053, 0.047]
Majority Race	-0.045	0.030	-1.468	0.142	[-0.104, 0.015]
SES	0.067	0.012	5.405	< 0.001	[0.043, 0.091]
Income	-0.031	0.016	-1.950	0.051	[-0.063, 0.000]
Platform: MTurk	0.225	0.083	2.706	0.007	[0.062, 0.389]
Platform: Toloka	0.382	0.064	5.940	< 0.001	[0.256, 0.509]
Platform: BeSample	0.439	0.070	6.236	< 0.001	[0.301, 0.577]
Platform: Aishiyan	0.081	0.161	0.501	0.619	[-0.234, 0.395]
Platform: Lancers	0.134	0.157	0.852	0.399	[-0.174, 0.441]
Phone	0.049	0.025	1.937	0.053	[-0.001, 0.099]
Completed the Whole Task x WEIRD, Linear	-0.145	0.049	-2.926	0.003	[-0.242, -0.048]
Completed the Whole Task x WEIRD, Quadratic	-0.113	0.044	-2.568	0.010	[-0.199, -0.027]

The regression output for the mixed-effects regression model on agreement with the item "Working for longer than 5 minutes could encourage the researchers to give me more work in the future" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous

variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: “Less than $5,000;” “$5,000 to $9,999;” "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled.

Appendix Table S24: **Study 5a, Task Enjoyment Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.354	0.074	-4.797	< 0.001	[-0.498, -0.209]
WEIRD, Quadratic	-0.064	0.049	-1.304	0.199	[-0.159, 0.032]
Completed the Whole Task	0.476	0.028	16.819	< 0.001	[0.420, 0.531]
Male	-0.233	0.025	-9.331	< 0.001	[-0.282, -0.184]
Age	0.004	0.015	0.261	0.794	[-0.025, 0.033]
Employed	0.068	0.027	2.488	0.013	[0.014, 0.121]
Student	-0.014	0.031	-0.462	0.644	[-0.075, 0.046]
Education	-0.066	0.013	-5.239	< 0.001	[-0.091, -0.042]
Urban	-0.016	0.027	-0.583	0.560	[-0.068, 0.037]
Majority Race	-0.036	0.032	-1.102	0.271	[-0.099, 0.028]
SES	0.018	0.013	1.417	0.157	[-0.007, 0.044]
Income	-0.043	0.016	-2.613	0.009	[-0.075, -0.011]
Platform: Toloka	0.196	0.075	2.626	0.010	[0.050, 0.342]
Platform: BeSample	0.267	0.076	3.500	0.001	[0.118, 0.417]
Platform: Aishiyan	-0.834	0.174	-4.807	< 0.001	[-1.175, -0.494]
Platform: Lancers	-0.670	0.167	-4.015	< 0.001	[-0.997, -0.343]
Phone	-0.083	0.026	-3.142	0.002	[-0.135, -0.031]
Completed the Whole Task x WEIRD, Linear	0.078	0.050	1.547	0.122	[-0.021, 0.176]
Completed the Whole Task x WEIRD, Quadratic	0.118	0.046	2.559	0.011	[0.028, 0.208]

The regression output for the mixed-effects regression model on agreement with the item "I enjoyed completing the task" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from

0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled.

Appendix Table S25: **Study 5a, Perceived Researcher Expectations Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.014	0.062	-0.223	0.824	[-0.135, 0.107]
WEIRD, Quadratic	0.068	0.041	1.646	0.109	[-0.013, 0.148]
Completed the Whole Task	0.349	0.030	11.807	< 0.001	[0.291, 0.407]
Male	0.075	0.024	3.094	0.002	[0.028, 0.123]
Age	-0.029	0.014	-1.984	0.047	[-0.057, -0.000]
Employed	-0.002	0.026	-0.068	0.946	[-0.054, 0.050]
Student	0.051	0.031	1.670	0.095	[-0.009, 0.111]
Education	-0.026	0.013	-2.029	0.042	[-0.050, -0.001]
Urban	-0.003	0.026	-0.105	0.916	[-0.055, 0.049]
Majority Race	-0.025	0.031	-0.790	0.430	[-0.086, 0.037]
SES	0.040	0.013	3.133	0.002	[0.015, 0.065]
Income	-0.012	0.017	-0.741	0.459	[-0.045, 0.020]
Platform: MTurk	0.334	0.085	3.915	< 0.001	[0.167, 0.502]
Platform: Toloka	0.259	0.062	4.150	< 0.001	[0.137, 0.382]
Platform: BeSample	0.317	0.069	4.572	< 0.001	[0.181, 0.453]
Platform: Aishiyan	0.805	0.151	5.346	< 0.001	[0.510, 1.100]
Platform: Lancers	0.536	0.147	3.654	0.001	[0.249, 0.824]
Completed the Whole Task x WEIRD, Linear	-0.150	0.054	-2.779	0.008	[-0.256, -0.044]
Completed the Whole Task x WEIRD, Quadratic	-0.137	0.048	-2.840	0.007	[-0.231, -0.042]

The regression output for the mixed-effects regression model on agreement with the item "The researchers wanted me to work for longer than 5 minutes" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The predictors include whether participants completed the whole task (Completed the Whole Task or Did Not Complete the Whole Task); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective

socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled.

Appendix Table S26: **Study 5a, The Effect of Self-Reported Motivation on Task Completion Among "WEIRD," "Partly (Sometimes) WEIRD," and "non-WEIRD" Countries.**

	b	*SE*	*z*	*P*	95% CI
WEIRD, Linear	-0.862	0.214	-4.021	< 0.001	[-1.282, -0.442]
WEIRD, Quadratic	-0.125	0.122	-1.026	0.305	[-0.364, 0.114]
Primary Needs	0.127	0.047	2.722	0.006	[0.035, 0.218]
Secondary Needs	-0.064	0.048	-1.326	0.185	[-0.158, 0.030]
Intrinsic	0.315	0.049	6.499	< 0.001	[0.220, 0.410]
Phone	-0.957	0.079	-12.139	< 0.001	[-1.111, -0.802]
Platform: MTurk	-0.491	0.702	-0.700	0.484	[-1.867, 0.884]
Platform: Toloka	0.311	0.236	1.318	0.187	[-0.151, 0.773]
Platform: BeSample	0.713	0.251	2.845	0.004	[0.222, 1.205]
Platform: Aishiyan	0.676	0.402	1.680	0.093	[-0.113, 1.464]
Platform: Lancers	0.016	0.379	0.043	0.965	[-0.726, 0.758]
Male	0.253	0.076	3.325	0.001	[0.104, 0.402]
Age	0.031	0.046	0.686	0.493	[-0.058, 0.121]
Employed	0.093	0.081	1.143	0.253	[-0.066, 0.251]
Student	0.017	0.093	0.186	0.852	[-0.164, 0.199]
Education	-0.034	0.039	-0.877	0.380	[-0.111, 0.042]
Urban	-0.018	0.080	-0.224	0.823	[-0.175, 0.139]
Majority Race	-0.293	0.096	-3.044	0.002	[-0.481, -0.104]
SES	0.001	0.039	0.036	0.971	[-0.075, 0.078]
Income	0.064	0.052	1.237	0.216	[-0.037, 0.165]
WEIRD, Linear x Primary Needs	0.099	0.081	1.216	0.224	[-0.061, 0.258]
WEIRD, Quadratic x Primary Needs	0.166	0.076	2.199	0.028	[0.018, 0.315]
WEIRD, Linear x Secondary Needs	-0.045	0.084	-0.531	0.596	[-0.210, 0.120]
WEIRD, Quadratic x Secondary Needs	-0.120	0.080	-1.507	0.132	[-0.277, 0.036]
WEIRD, Linear x Intrinsic	-0.027	0.087	-0.308	0.758	[-0.198, 0.144]
WEIRD, Quadratic x Intrinsic	0.084	0.078	1.068	0.285	[-0.070, 0.238]

The regression output for the mixed-effects logistic regression model on task completion in Study 5a among participants who took the additional platform motivation measures (N = 4,368 independent participants recruited on seven crowd-sourcing platforms). The predictors include Condition (Explicitly No Pay or Ambiguous Pay); "WEIRD, Linear" and "WEIRD, Quadratic," which represent the linear and quadratic terms of the "WEIRD" country classification—an ordinal variable. The choices answer for the motivations type categories were "To earn money for food, rent, or bills," which we renamed "Primary Needs;" "To earn money for hobbies, gadgets, or going out," which we renamed "Secondary Needs;" and five answer choices that we merged into a category of "Intrinsic" reasons: "To kill time;" "Because the tasks are fun;" "To benefit society," "To learn new things," and "To share my opinions with someone who cares about what I think." The "WEIRD" classification is adapted from Talhelm et al (under review). Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Education is a categorical variable for the highest level of education workers have obtained or are in the process of obtaining. Income is a continuous variable that ranges between 5,000 and 150,000, taken as a midpoint from reported annual income (in USD) with the following answer choices: "Less than $5,000;" "$5,000 to $9,999;" "$10,000 to $14,999;" "$15,000 to $19,999;" "$20,000 to $29,999;" "$30,000 to $39,999;" "$40,000 to $49,999;" "$50,000 to $59,999;" "$60,000 to $69,999;" "$70,000 to $79,999;" "$80,000 to $99,999;" "$100,000 to $149,999;" and "$150,000 or more." SES is a continuous ("ladder") variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Majority Race is a binary variable indicating whether participants belonged to the majority race reported by workers in their respective countries. Platform variables (MTurk, Toloka, BeSample, Aishiyan, Lancers) are binary variables indicating the platform on which participants were recruited. Phone is a binary variable indicating whether the participant used a phone (versus a laptop) to complete the task. The model includes a random intercept for country. All non-binary variables are scaled.

CONDITIONS FROM DELLAVIGNA AND POPE (2018) INCLUDED IN AND EXCLUDED FROM STUDY 1

Exclusion criteria are summarized in more detail in the Methods section of Study 1. To reiterate, we categorized as "monetary" pay-for-effort treatments with linear, immediate, and guaranteed piece-rates. Conditions in which the participants could not earn any additional money for themselves—either immediately or at some point in the future—were categorized as "non-monetary." Below we report every condition from the previous study (DellaVigna & Pope, 2018), for which the data were originally collected, and whether these data were included in our analysis in Study 1.

Appendix Table S27a: **Monetary Versus Non-Monetary Treatments from DellaVigna and Pope (2018) Included in Study 1.**

Monetary Treatments	Non-Monetary Treatments
Self–1 cent for 1,000 points "You will be paid an extra 1 cent for every 1,000 points."	*Flat-Fee Control* "Your score will not affect your payment."
Self–1 cent for 100 points "You will be paid an extra 1 cent for every 100 points."	*Unconditional Gift* "In appreciation for performing this task, you will be paid a bonus of 40 cents. Your score will not affect your payment."
Self–4 cents for 100 points "You will be paid an extra 4 cents for every 100 points."	*Please Try* "Please try as hard as you can."
Self–10 cents for 100 points "You will be paid an extra 10 cents for every 100 points."	*Social Comparison* "We will show you how well you did relative to others."
	Social Norm "Many participants scored more than 2,000."
	Charity–1 cent for 100 points "The Red Cross will be given 1 cent for every 100 points."
	Charity–10 cents for 100 points "The Red Cross will be given 10 cents for every 100 points."

Incentive treatments from a previous study (DellaVigna & Pope, 2018) included in Study 1 and categorized as either monetary or non-monetary.

Appendix Table S27b: **Treatments from DellaVigna and Pope (2018) Excluded from Study 1 and Reasons for Exclusions.**

Treatment	Reason
Probability Weighting–Low Chance, High Reward "You will have a 1% chance of an extra $1 for every 100 points."	Extra payment is not guaranteed
Probability Weighting–High Chance, Low Reward "You will have a 50% chance of an extra 2 cents for every 100 points."	Extra payment is not guaranteed
Delay Discounting–2 weeks "You will be paid an extra 1 cent for every 100 points (2-week delay)."	Extra payment is not immediate
Delay Discounting–4 weeks "You will be paid an extra 1 cent for every 100 points (4-week delay)."	Extra payment is not immediate
Gain–40 cents "You will be paid an extra 40 cents if you score at least 2,000 points."	Extra payment is not linear
Loss–40 cents "You will be paid an extra 40 cents. However, you will lose this bonus unless you score."	Extra payment is not linear
Gain–80 cents "You will be paid an extra 80 cents if you score at least 2,000 points."	Extra payment is not linear

Incentive treatments from a previous study (DellaVigna & Pope, 2018) excluded from Study 1, together with the reasons for each exclusion (i.e., which criteria for being classified as "monetary" were not met).

ATTENTION AND COMPREHENSION CHECKS

Attention Check

Participants across studies first read the following two paragraphs:

"We are developing a machine learning image classification database, and we need your help! You are about to see a collection of images one by one, and your task is to assess whether each image contains a building or not. If you see that the image contains a building, answer YES. If you see the image does not contain a building, answer NO. If any image doesn't load, please answer NO.

Every ten images, you will be given an opportunity to be finished with the task. You can view as many images as you like. After the first 10 images completed, you will get the fee for your participation regardless of how many images you view or when you choose to quit the task."

Then, participants saw the first attention check that asked them to select the purpose of the task ("developing a machine-learning image-classification database") out of the three multiple choice options. Those who didn't pass the check were not allowed to proceed with the survey.

Comprehension Check

After being assigned to a condition and learning about the pay structure and their incentive treatment assignment, participants were asked, "How much extra pay will you get for every 10 images?" The four multiple choice answers ranged from "I will not receive any additional pay based on the number of images I complete" to "20 cents per 10 images."

Participants had two attempts to answer the questions correctly. Those who failed on both attempts were not allowed to proceed to the image-rating task.

IMAGES FROM THE EXPERIMENTAL TASK

Appendix Image S1: Example of an Image from the Image-Rating Task.

Example of an image (Jose and Roxanne, 2010) included in the image-rating task that did not contain a building.

Appendix Image S2: **Example of an Image from the Image-Rating Task.**

Example of an image (MasterPhoto-DK, 2007) included in the image-rating task that contained a building.

Appendix Image S3: **Example of a Data Entry Sheet Used in Studies 5A and 5B.**

Client Number	Number of Orders	Month of First Order	Date of Last Order	Contacted Client Support	Left a Review
17829	2	3/2022	8/2023	784032	0
23950	4	10/2022	1/2023	245273	1
21990	3	2/2018	1/2022	594315	1
18892	14	3/2023	4/2023	528463	1
19203	1	2/2023	4/2023	955434	1
48739	8	7/2018	7/2019	605563	0
43633	10	9/2023	2/2023	807079	0
40416	5	10/2016	6/2017	363722	0
36513	15	9/2015	7/2016	306844	1
53656	4	5/2023	6/2023	744547	0
25821	6	7/2020	11/2020	490760	1
11976	4	5/2021	5/2022	943885	0
17547	5	4/2023	5/2023	962859	0
88703	10	1/2023	7/2023	229559	0
24086	15	7/2022	10/2022	930338	1
54191	3	12/2021	7/2022	313890	1
17451	8	11/2019	1/2020	473582	0
73615	12	4/2020	12/2020	302708	1
71417	7	8/2018	12/2018	929854	1
23935	4	4/2017	11/2019	453745	0
69776	5	9/2021	1/2023	739488	1
47187	9	2/2022	3/2023	467449	0

Example of a data entry sheet that participants worked on in Studies 5a and 5b.

MAIN EFFECTS FOR REPORTED REGRESSIONS IN STUDIES 1–4

Below we report the main effects for the regressions on the mean number of images rated in Studies 1–4 in the monetary and (pooled) non-monetary conditions, controlling for age, gender, and education.

Study 1

We found both a main effect of treatment type, $b = 159.85$, $t(6287) = 2.91$, $P = 0.004$, 95% CI 52.35 to 267.35, and country, $b = 195.69$, $t(6287) = 5.43$, $P < 0.001$, 95% CI 125.05 to 266.34 (Fig. 1 and Table 1).

Study 2A

The main effect of incentive type for the regression on the number of images rated with incentive type, country, age, gender, and education as predictors was statistically significant, $b = 8.79$, $t(2145) = 3.11$, $P = 0.002$, 95% CI 3.25 to 14.33; the main effective of culture, $b = 2.18$, $t(2145) = 0.79$, $P = 0.429$, 95% CI -3.22 to 7.58 was not (Fig. 2 and Table 2).

Study 2B

Participants across samples rated more images in the monetary condition, $b = 27.49$, $t(2143) = 9.42$, $P < 0.001$, 95% CI 21.76 to 33.21 for the comparison between Mexico and the US sample with the same nominal pay, and $b = 27.49$, $t(2167) = 9.31$, $P < 0.001$, 95% CI 21.70 to 33.28 for the comparison between Mexico and the same subjective pay sample in the US (Fig. 3 and Table 3). Additionally, participants in the US completed fewer images than did participants in Mexico: Mexico and the US sample with the same nominal pay: $b = -37.51$, $t(2143) = -14.32$, $P < 0.001$, 95% CI -42.64 to -32.37; Mexico and the US sample with the same subjective pay: $b = -36.21$, $t(2167) = -13.87$, $P < 0.001$, 95% CI -41.33 to -31.09.

Study 2C

Participants in South Africa and the US completed more images in the monetary than non-monetary conditions, $b = 25.29$, $t(1303) = 8.78$, $P < 0.001$, 95% CI 19.64 to 30.93 for South Africa and the US sample with the same nominal pay; $b = 25.22$, $t(1303) = 8.63$, $P < 0.001$, 95% CI 19.48 to 30.95 for South Africa and the US sample with the same subjective pay (Fig. 4 and Table 4). Participants in the US completed fewer images than did participants in South Africa, $b = -8.97$ (South Africa and the US sample with the same nominal pay); $t(1303) = -2.97$, $P = 0.003$, 95% CI -14.90 to -3.04; $b = -10.78$, $t(1303) = -3.57$, $P < 0.001$, 95% CI -16.70 to -4.86 (South Africa and the US sample with the same subjective pay).

Study 3A

We found a statistically significant main effect of country, with participants in India completing more image-ratings than did participants in the US, $b = -26.03$, $t(726) = -5.71$, $P < 0.001$, 95% CI -34.99 to -17.08. We did not observe a statistically significant effect of incentive, $b = 0.83$, $t(726) = 0.19$, $P = 0.847$, 95% CI -7.57 to 9.22 (Fig. 5 and Table 5).

Study 4

A multiple linear regression on the mean number of images rated with incentive and language as predictors showed a statistically significant main effect of incentive, $b = 12.63$, $t(2061) = 5.56$, $P < 0.001$, 95% CI 8.17 to 17.09, but not language, $b = -1.74$, $t(2061) = -0.76$, $P = 0.445$, 95% CI -6.22 to 2.74 (Fig. 6 and Table 6).

EXCLUSION DATA AND CRITERIA FOR STUDIES 2–5

Among the participants who passed the attention checks, the only criteria for inclusion were that participants should (1) receive the payment code, which meant they got to the end of the survey, (2) rate at least one image, and (3) not complete the same survey more than once. Completing the task could be achieved in one of the two ways: (1) completing the whole 10 minutes of the task and timing out or (2) choosing to quit the task at one of the screens that appeared after every 10 image-ratings and asked whether participants wanted to continue to rate images. For criterion (3), we used Worker IDs to identify duplicate submissions for the same survey.

Exclusion Data

Study 2A

China

For the survey in China, we received a total of 2,029 unique submissions. Of these, 9 participants did not provide consent to participate in the study; 702 people did not pass the attention check (N = 500) or comprehension check (N = 168) or willingly quit the survey before getting to the image-rating task (N = 34). Additionally, 222 participants willingly quit the survey during the task that contained DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

Out of the people who received the payment code (N = 1,096), which means they completed the whole survey, we removed the participants who did not rate any images (N = 5) or reported being younger than 18 years old (N = 5). Every remaining participant completed at least one image, lived in China, and was at least 18 years old. This left us with 1,086 participants.

In China, we did not have a variable similar to one's "Worker ID." Although we did collect participants' AliPay accounts, we thought it was unreasonable to automatically classify repeating AliPay accounts as duplicates: not everyone has their own AliPay account, and it is relatively common for people to receive payments on someone else's behalf.

UK

For the survey in the UK, we received a total of 1,164 unique submissions. From these submissions, we excluded those where participants did not provide consent to participate in the study (N = 1). Further, 90 people failed the attention (N = 18) or comprehension (N = 72) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 0).

From those who received the payment code (N = 1,073), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 6). Every remaining participant completed at least one image, lived in the UK, and was at least 18 years old. This left us with 1,067 participants.

Study 2B

Mexico

For the survey in Mexico, we received a total of 1,268 unique submissions. Of these, 2 participants did not give consent to participate in the study. Further, 156 people failed the attention (N = 60) or comprehension (N = 94) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 2). Additionally, 5 participants willingly quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 1,105), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by

their input of the Worker IDs (N = 52). Every remaining participant completed at least one image, lived in Mexico, and was at least 18 years old. This left us with 1,053 participants.

US Same Nominal Pay

For the US survey with the same nominal pay, we received a total of 1,195 unique submissions. Of these, 79 people failed the attention (N = 43) or comprehension check (N = 35) or willingly quit the survey before getting to the image-rating task that contained the DV (N = 1). Additionally, 5 participants willingly quit the survey during the task that contained DV (not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 1,111), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 13). Every remaining participant completed at least one image, lived in the US, and was at least 18 years old. This left us with 1,098 participants.

US Same Subjective Pay

For the US survey with the same subjective pay, we received a total of 1,219 unique submissions. Of these people, 1 participant did not give consent to participate in the study. Seventy-five failed the attention (N = 46) or comprehension (N = 29) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 0). Additionally, 11 participants willingly quit the survey during the task that contained DV (not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 1,132), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by

their input of the Worker IDs (N = 10). Every remaining participant completed at least one image, lived in the US, and was at least 18 years old. This left us with 1,122 participants.

Study 2C

South Africa

For the survey in South Africa, we received a total of 717 unique submissions. Of these, 4 participants did not provide consent to participate in the study. Fifty-one either failed the attention (N = 10) or comprehension (N = 41) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 0). Additionally, 6 participants willingly quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 656), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 4). Finally, we excluded participants who did not rate a single image (N = 1) and those who stated that they did not currently live in South Africa (N = 2). Every remaining participant completed at least one image, currently lived in South Africa, and was at least 18 years old. This left us with 649 participants.

US Same Nominal Pay

For the US survey with the same nominal pay, we received a total of 700 unique submissions. Of these, 23 failed the attention (N = 8) or comprehension (N = 13) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 2). Additionally, 14 participants willingly quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 663), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their Worker IDs (N = 1). Every remaining participant completed at least one image, currently lived in the US, and was at least 18 years old. This left us with 662 participants.

US Same Subjective Pay

For the US survey with the same subjective pay, we received a total of 697 unique submissions. Of these, 1 participant did not provide consent to participate in the study. Twenty-four failed the attention (N = 5) or comprehension (N = 18) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 1). Additionally, 6 participants willingly quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 666), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 2) and participants who stated that they did not currently living in the US (N = 2). Every remaining participant completed at least one image, currently lived in the US, and was at least 18 years old. This left us with 662 participants.

Study 3A

India

For the survey in India, we received a total of 602 unique submissions. Of these, 2 participants did not currently live in India and 10 did not give consent to participate in the study. Further, 160 participants failed the attention (N = 62) or comprehension check (N = 97) or willingly quit the survey before rating the image (N = 1). Additionally, 17 participants willingly

quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 413), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 61). Every remaining participant completed at least one image, currently lived in India, and was at least 18 years old. This left us with 352 participants.

US

For the survey in the US, we received a total of 432 unique submissions. Of these, 4 participants did not give consent to participate in the study, 35 people either did not pass the attention (N = 16) or comprehension (N = 18) check or willingly quit the survey before rating a single image (N = 1). Additionally, 7 participants willingly quit the survey during the task that contained the DV (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task.

From those who received the payment code (N = 386), which means they completed the whole survey, we removed the second or further attempts by the same participant as identified by their input of the Worker IDs (N = 4). Every remaining participant completed at least one image, currently lived in the US, and was at least 18 years old. This left us with 382 participants.

Study 3B

For the US survey, we received a total of 566 unique submissions. Of these, we removed 23 people who failed the attention (N = 7) or comprehension (N = 15) check or willingly quit the survey before getting to the image-rating task that contained the DV (N = 1). Additionally, 5 participants willingly exited the survey during the task that contained the dependent variable and thus did not complete the survey.

From those who completed the whole survey and received the payment code (N = 538), we removed 1 participant who stated that they did not currently live in the US. Every remaining participant completed at least one image, currently lived in the US, and was at least 18 years old. This left us with 537 participants.

Study 4

For the survey in India, we received a total of 6,246 responses. Of these, 825 were excluded as they were identified as "Spam" by the (internal) Qualtrics quality-control mechanism. Of the remaining 5,421 unique submissions, 2,386 were removed as they were identified by either duplicate phone numbers or email addresses as second or further attempts by the same participant. This number is higher than in the previous studies, because participants were recruited on Facebook and not a crowd-sourcing website, which means that nothing prevented participants from using the survey link again, if they wanted to do so.

Of the remaining 3,035 participants, 933 people either failed the attention (N = 589) or comprehension (N = 339) check or willingly quit the survey before the image-rating task (N = 5). Additionally, 33 participants willingly quit the survey during the image-rating task (i.e., not during one of the screens that allowed them to do so without penalty) and thus did not complete the task. Finally, 2 participants completed the survey without rating any images, and 2 participants provided phone number input that was not numeric. Every remaining participant completed at least one image, currently lived in India, and was at least 18 years old. This left us with 2,065 participants.

Studies 5A and 5B

For Study 5A, we received data from a total of 20,761 participants who completed the data entry task. We first removed 1,327 duplicate entries, which were identified through repeated Worker IDs. Subsequently, 128 participants were excluded for being under the age of 18, and 858 entries were removed because they listed different countries for nationality and current residence.

Furthermore, 1,919 participants failed an attention check that asked them to select the number “2” on a scale, and an additional 1,044 participants were excluded because they misunderstood the instructions and believed they would be paid for their participation. After these exclusions, we were left with a dataset of 15,485 participants who met the study's criteria: 7,823 participants in Study 5A and 7,578 participants in Study 5B.

Note Regarding Samples in Studies 2A and 3A

Some participants completed the task and received the payment code but did not press the button that officially ended the survey. We downloaded the data for Studies 2a and 3a before the responses from these participants were counted as completed by Qualtrics. On Qualtrics, such responses do not appear in the data for one week. In Study 2a, 13 participants in China and 62 participants in the UK were not included in the original dataset we used for the analyses. In Study 3a, the number of these participants was 7 in India and 1 in the US.

We only noticed the data from these participants after we had downloaded the raw Qualtrics data during the revision process to upload the raw Qualtrics files onto OSF together with the rest of the study materials. We chose not to update the results in the paper to reflect these additional participants. However, the main results reported in the paper do not change regardless of whether these "late" participants are included or excluded.

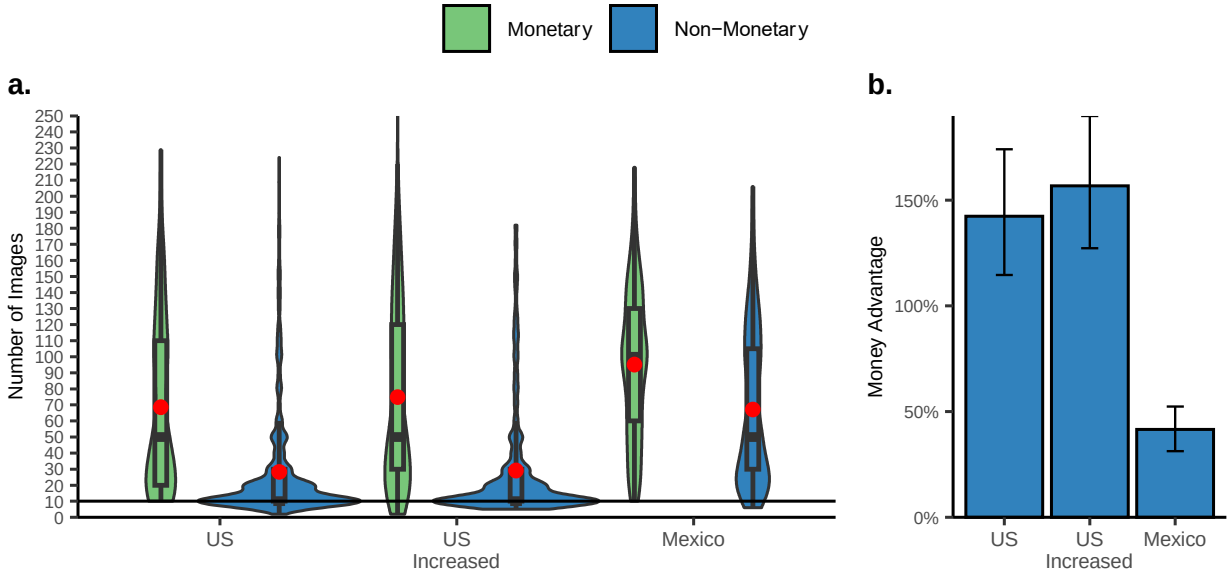
Thus, the central interaction between country and incentive type in Study 2a reported in the main text as $b = 38.40$, $t(2145) = 9.65$, $P < 0.001$, 95% CI 30.59 to 46.20 becomes $b = 38.85$, $t(2214) = 9.92$, $P < 0.001$, 95% CI 31.18 to 46.53, if these "late" participants are included. The analogous interaction in Study 3a, reported as $b = 13.77$, $t(726) = 2.31$, $P = 0.021$, 95% CI 2.09 to 25.45, becomes $b = 14.16$, $t(734) = 2.39$, $P = 0.017$, 95% CI 2.52 to 25.80.

Note Regarding the US Samples in Study 2B

After collecting the data in the US, we noticed that some participants completed both the survey with the same nominal pay and the survey with increased pay, as we ran the two surveys concurrently and could not automatically exclude participants from taking both surveys. In the main text, attempts for both surveys are included. Here, we report the key results if we only include the participants' first attempt across the two surveys.

Thus, if a participant completed the survey with the nominal pay first, we would keep this response and exclude their data from the sample with the same subjective pay. Performing these exclusions would leave us with 1,048 participants in the sample with the same nominal pay and 953 participants in the sample with the same subjective pay. None of the results reported in the main text significantly change based on whether we include or exclude these participants. The interaction between country and incentive type, controlling for age, gender, and education, reported in the main text remains statistically significant for the comparison between Mexico and the US sample with the same nominal pay, $b = 13.10$, $t(2093) = 3.20$, $P = 0.001$, 95% CI 5.07 to 21.14, as well as the US sample with increased pay, $b = 17.86$, $t(1998) = 4.16$, $P < 0.001$, 95% CI 9.44 to 26.27 (Fig. S1).

Fig. S1: **Study 2b, Monetary Versus Pooled Non-Monetary Conditions in Mexico and the US (Duplicate Participants in the US Excluded).**



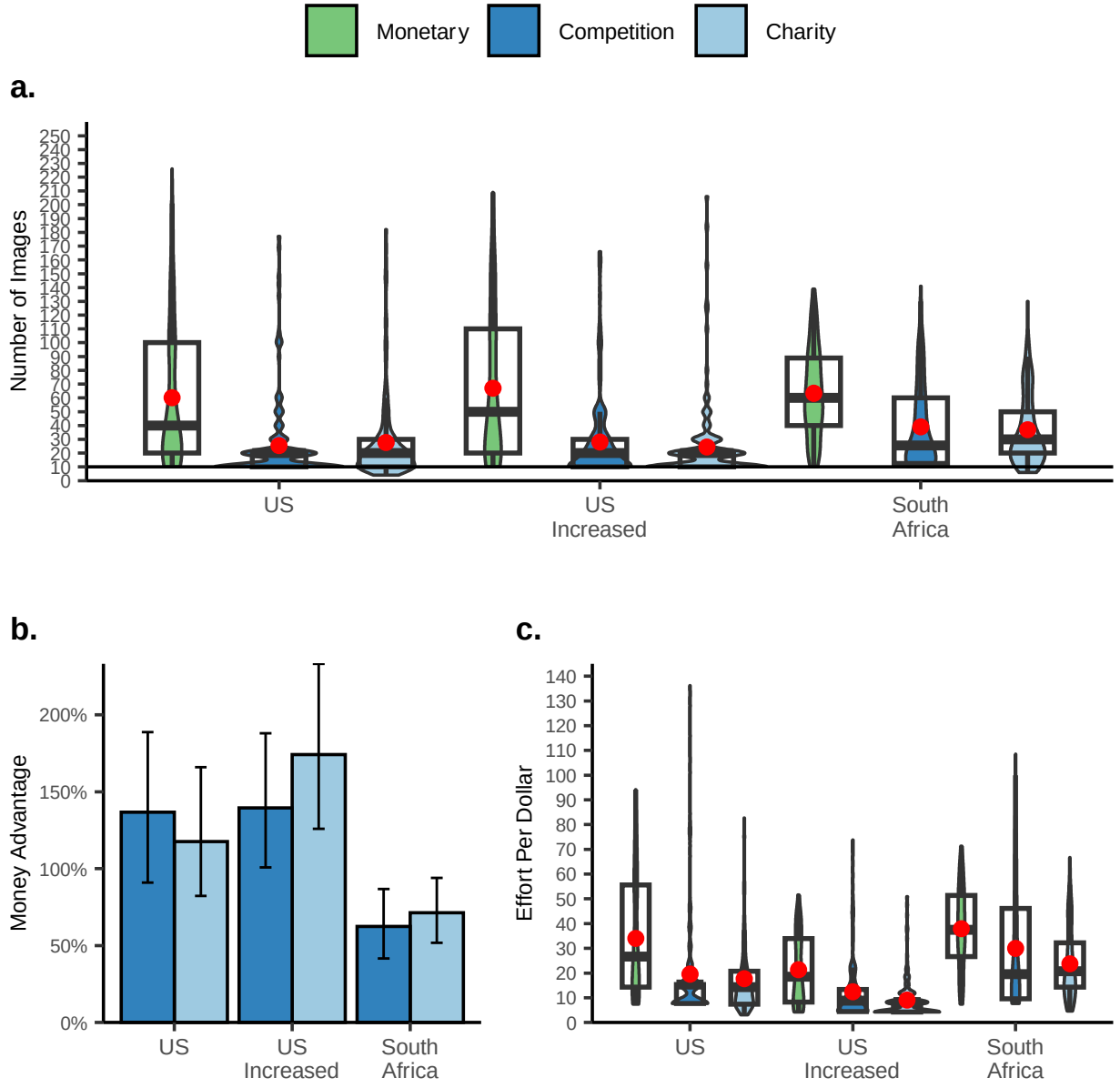
Effects of a monetary incentive (green) and pooled non-monetary treatments (flat fee and social norm; blue) in Mexico (N = 1,053 participants recruited on Prolific) and two samples in the US: one with the same nominal pay (N = 1,048 participants recruited on Prolific) as in Mexico and one with the same subjective (Becker et al., 1964) pay (N = 953 participants recruited on Prolific) as in Mexico. Second attempts from the participants in the US who completed both versions of the survey (N_{Nominal} = 50 participants, $N_{\text{Subjective}}$ = 169 participants) are excluded. Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between country and incentive type in a multiple linear regression model *is* statistically significant for the comparison between Mexico and the US sample with the same nominal pay, $b = 13.10$, $t(2093) = 3.20$, $P = 0.001$, 95% CI 5.07 to 21.14, as well as the US sample with the same subjective pay, $b = 17.86$, $t(1998) = 4.16$, $P < 0.001$, 95% CI 9.44 to 26.27. Panel B shows the money advantage—that is, how much more effective money is than the pooled non-monetary treatments in each sample. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images rated in the monetary versus pooled non-monetary conditions.

MEAN EFFORT AND COST-EFFECTIVENESS IN INDIVIDUAL CONDITIONS IN STUDY 2C

Compared to the monetary incentive condition, the competition condition decreased effort more in the US than in South Africa, both in the same nominal pay sample in the US, b = -9.85, $t(1301)$ = -1.98, P = 0.048, 95% CI -19.60 to -0.11 and the sample with the same subjective pay, b = -14.51, $t(1301)$ = -2.90, P = 0.004, 95% CI -24.33 to -4.69. For the charity condition, the interaction between culture and incentive type was statistically significant for the same-subjective-pay sample, b = -16.38, $t(1301)$ = -3.21, P = 0.001, 95% CI -26.38 to -6.38, but not for the same-nominal-pay sample, b = -6.09, $t(1301)$ = -1.22, P = 0.222, 95% CI -15.86 to 3.69 (Fig. S2a and Appendix Table S10).

In South Africa, the monetary incentive was significantly more cost-effective than the competition condition, two-sided Welch's $t(236.70) = 3.81$, $P < 0.001$, $P_{Bonf} = 0.001$, $Mean_{\text{difference}} = 7.89$, $d = 0.39$, 95 CI% 3.81 to 11.97, and the charity condition, two-sided Welch's $t(331.79) = 9.58$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 14.12$, $d = 0.91$, 95 CI% 11.22 to 17.02 (Fig. S2c). In both US samples, money was similarly more cost-effective than the competition condition (Same Nominal Pay: two-sided Welch's $t(337.20) = 6.48$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 14.44$, $d = 0.62$, 95 CI% 10.06 to 18.83; Increased Pay: two-sided Welch's $t(358.34) = 7.05$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 8.88$, $d = 0.66$, 95 CI% 6.40 to 11.36) and the charity condition (Same Nominal Pay: two-sided Welch's $t(468.45) = 9.21$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 16.32$, $d = 0.81$, 95 CI% 12.84 to 19.80; Increased Pay: two-sided Welch's $t(474.17) = 12.34$, $P < 0.001$, $P_{Bonf} < 0.001$, $Mean_{\text{difference}} = 12.22$, $d = 1.08$, 95 CI% 10.28 to 14.17).

Fig. S2: **Study 2c, Monetary Versus Individual Non-Monetary Conditions in the US and South Africa.**



Effects of the three individual conditions: monetary (green), competition (dark blue), and charity (light blue) in South Africa (N = 649 participants recruited on Prolific) and two samples in the US: one with the same nominal pay (N = 662 participants recruited on Prolific) as in South Africa and the other with the same subjective (Becker et al., 1964) pay (N = 662 participants recruited on Prolific) as in South Africa. Panel A shows the central tendency and distribution of effort by incentive type and country. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. The interaction between culture and the competition condition is significant for the sample with the US sample with the same nominal pay, b = -9.85, $t(1301)$ = -1.98, P = 0.048, 95% CI -19.60 to -0.11, and the sample with the same subjective pay, b = -14.51, $t(1301)$ = -2.90, P = 0.004, 95% CI -24.33 to -4.69. The interaction between culture and

the charity condition in a multiple linear regression model *i*s statistically significant for the comparison between South Africa and the same subjective-pay US sample, $b = -16.38$, $t(1301) = -3.21$, $P = 0.001$, 95% CI -26.38 to -6.38, but not the same-nominal-pay sample, $b = -6.09$, $t(1301) = -1.22$, $P = 0.222$, 95% CI -15.86 to 3.69. Panel B shows the money advantage—that is, how much more effective money is than each of the two non-monetary treatments in each sample. Panel C shows the central tendency and distribution of the cost-effectiveness (effort per dollar spent) by incentive type and country. Graph elements are analogous to those in Panel A, with the width of each violin corresponding to the frequency of observations at any given level of cost-effectiveness (effort per dollar spent) rated on the y-axis. The results for cost-effectiveness are summarized in the Appendix. In Panel B, error bars are bootstrapped 95% CIs for the mean relative difference in the number of images in the monetary and each of the non-monetary conditions.

Appendix Table S28: **Study 2c, Monetary Versus Individual Non-Monetary Conditions in the US and South Africa.**

	b	*SE*	*t*	*P*	95% CI
South Africa and the US (Same Nominal Pay)					
Competition	-24.32	3.52	-6.92	< 0.001	[-31.21, -17.42]
Charity	-26.29	3.57	-7.37	< 0.001	[-33.29, -19.30]
US	-0.99	2.96	-0.33	0.738	[-6.81, 4.82]
Female	0.47	2.07	0.22	0.822	[-3.60, 4.53]
Non-Binary	-6.79	7.42	-0.92	0.360	[-21.34, 7.76]
Education	-6.95	2.08	-3.35	0.001	[-11.03, -2.88]
Age	-3.12	1.16	-2.69	0.007	[-5.40 -0.84]
Competition × US	-9.85	4.97	-1.98	0.048	[-19.60, -0.11]
Charity x US	-6.09	4.98	-1.22	0.222	[-15.86, 3.69]
South Africa and the US (Same Subjective Pay)					
Competition	-24.22	3.57	-6.79	< 0.001	[-31.22, -17.22]
Charity	-26.25	3.62	-7.25	< 0.001	[-33.35, -19.15]
US	4.64	3.02	1.53	0.125	[-1.29, 10.57]
Female	-2.13	2.09	-1.02	0.309	[-6.23, 1.98]
Non-Binary	-12.97	7.92	-1.64	0.102	[-28.50, 2.57]
Education	-4.27	2.11	-2.02	0.044	[-8.42, -0.12]
Age	-1.11	1.19	-0.93	0.351	[-3.43, 1.22]
Competition × US	-14.51	5.01	-2.90	0.004	[-24.33, -4.69]
Charity x US	-16.38	5.10	-3.21	0.001	[-26.38, -6.38]

Regression output for the multiple linear regression on the number of images rated in the individual conditions in the Study 2c sample from South Africa (N = 649 recruited on Prolific) and each of the two Study 2c samples in the US: one with the same nominal pay (N = 662 recruited on Prolific) as in South Africa and the other with the same subjective (Becker et al., 1964) pay (N = 662 recruited on Prolific) as in South Africa. Competition is a binary variable for whether participants were assigned to the competition condition. Charity is a binary variable for whether participants were assigned to the charity condition. US is a binary variable for whether participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Age is a continuous variable and is scaled.

PROBABILITY OF QUITTING THE TASK AT THE FIRST OPPORTUNITY IN STUDIES 2A–4

As an additional measure of effort, we looked at the proportion of participants who chose to quit the task without losing the base pay for participation when provided with the first opportunity to do so. Everyone in our study was explicitly instructed that they could quit the task without losing the base pay for participation after rating only 10 images.

Study 2A

Unsurprisingly, few people regardless of culture quit the task immediately when presented with a pay-for-effort incentive (8.0% in China vs. 4.6% in the UK). However, in the non-monetary conditions, 34.3% of participants in the UK quit the task at the first opportunity, while only 14.2% of participants in China did so. When we conducted a logistic regression on the probability of quitting after rating exactly 10 images, we found a statistically significant interaction between incentive type and country, $b = -1.73$, $z = -5.03$, $P < 0.001$, OR = 0.18, 95% CI 0.09 to 0.35.

There was a statistically significant main effect of incentive type, with the participants quitting less frequently when offered a monetary bonus, $b = -0.66$, $z = -2.95$, $P = 0.003$, OR = 0.52, 95% CI 0.33 to 0.79, and country, with the UK participants being more likely to quit, $b = 1.30$, $z = 8.14$, $P < 0.001$, OR = 3.68, 95% CI 2.69 to 5.04.

Study 2B

Across samples, a relatively small proportion of participants quit the task immediately when presented with a pay-for-effort incentive, especially in Mexico (2.6% in Mexico; 15.1% in the US sample with the sample nominal pay and 13.3% in the US sample with the same subjective pay). However, when there was no monetary incentive to continue, 51.2% of US

participants in the same nominal pay sample and 54.5% in the increased pay sample exited the task at the first opportunity. In Mexico, 91.7% continued even when they could have quit early and received the same total pay for their work.

A logistic regression was used to analyze the role of incentive type and culture on the probability of quitting at the first opportunity without losing the base pay. Controlling for age, gender, and education, the interaction between incentive type and culture on the probability of quitting at 10 images was not statistically significant for the comparison between Mexico and the US sample with the same nominal pay, $b = -0.57$, $z = -1.41$, $P = 0.157$, OR = 0.57, 95% CI 0.27 to 1.32, but statistically significant for the US sample with the same subjective pay, $b = -0.85$, $z = -2.11$, $P = 0.034$, OR = 0.43, 95% CI 0.20 to 0.99.

Study 2C

In South Africa, 4.2% of participants quit at the first opportunity in the monetary condition. In the US, that number was 18.5% in the sample with the same nominal pay as in South Africa and 16.0% in the sample with the same subjective pay. Across the two non-monetary conditions, 21.8% of people in South Africa quit after rating 10 images and 43.8% and 45.5% of people in the two US samples, respectively.

When we conducted a logistic regression on the probability of quitting the task at the first opportunity, we found that participants were less likely to quit the task in the monetary condition, $b = -1.87$, $z = -6.11$, $P < 0.001$, OR = 0.15, 95% CI 0.08 to 0.27 for the comparison between South Africa and the US sample with the same nominal pay (and $b = -1.88$, $z = -6.13$, $P < 0.001$, OR = 0.15, 95% CI 0.08 to 0.27 for the comparison between South Africa and the US sample with the same subjective pay). Participants in the US were more likely to quit the task than participants in South Africa, $b = 1.10$, $z = 5.88$, $P < 0.001$, OR = 3.00, 95% CI 2.09 to 4.34

for the sample in the US with the same nominal pay (and $b = 1.17$, $z = 6.35$, $P < 0.001$, OR = 3.20, 95% CI 2.25 to 4.62 for the sample with the same subjective pay). The interaction was not statistically significant for the comparison between the South African sample and the US sample with the same nominal pay, $b = 0.59$, $z = 1.67$, $P = 0.094$, OR = 1.81, 95% CI 0.92 to 3.73, and between South Africa and the US sample with the same subjective pay, $b = 0.40$, $z = 1.11$, $P = 0.266$, OR = 1.49, 95% CI 0.75 to 3.09, showing that participants in the US were more likely to quit in both types of conditions and not in the non-monetary conditions in particular.

Study 3A

Unlike in all the previous studies, in Study 3a a higher proportion of people in the US quit at the first opportunity (35.4% in the minimal pay condition compared to 14.9% in India) in the monetary condition. In the norm condition, 47.4% of people in the US quit the task compared to 18.1% in India. When we ran a multiple linear regression and controlled for age, gender, and education, we found a statistically significant main effect of country, $b = 1.63$, $z = 6.11$, $P < 0.001$, OR = 5.09, 95% CI 3.05 to 8.68, showing that people in the US were more likely to quit the task after 10 images regardless of condition, but neither a main effect of incentive, $b = -0.27$, $z = -0.92$, $P = 0.359$, OR = 0.77, 95% CI 0.43 to 1.35, nor their interaction, $b = -0.24$, $z = -0.66$, $P = 0.510$, OR = 0.79, 95% CI 0.39 to 1.60.

Study 3B

A logistic regression on quitting after rating ten images showed that people in the minimal pay condition were less likely to quit at the first opportunity (46.5%) compared to people in the points condition (63.9%), $b = -0.71$, $z = -4.03$, $P < 0.001$, OR = 0.49, 95% CI 0.35 to 0.69.

Study 4

In the monetary condition, 7.9% of participants who completed the study in English and 9.5% who completed the study in Hindi exited the task after rating 10 images—the first time they could do so without losing their base pay. In the norm condition, those numbers were 24.9% in English and 25.3% in Hindi. A logistic linear regression on the likelihood of quitting at the first opportunity was not statistically significant, $b = -0.17$, $z = -0.66$, $P = 0.509$, OR = 0.84, 95% CI 0.50 to 1.41. The regression further showed a statistically significant main effect of incentive, $b = -1.18$, $z = -6.54$, $P < 0.001$, OR = 0.31, 95% CI 0.21 to 0.44, but not of language, $b = -0.02$, $z = -0.16$, $P = 0.871$, OR = 0.98, 95% CI 0.74 to 1.29.

DEMOGRAPHIC VARIABLES AND SAMPLE CHARACTERISTICS

Appendix Table S29a: **Demographic Variables for Study 1.**

	M	*SD*	*N*
US			
Monetary Incentive	36.7%		5526
Female	57.8%		5526
Education	50.6%		5526
Age Category	2.62	1.29	5526
India			
Monetary Incentive	35.8%		768
Female	32.2%		768
Education	84.0%		768
Age Category	2.41	1.02	768

Monetary Incentive is a binary variable for whether participants were in one of the monetary (pay-for-effort) conditions. US is a binary variable for whether participants were part of the US sample. Female is a categorical variable for gender. Age is measured on a scale, where 1 = "18–24"; 2 = "25–30"; 3 = "31–40"; 4 = "41–50"; 5 = "51–64"; 6 = "65 or higher." Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment.

Appendix Table S29b: **Demographic Variables for Studies 2a–c.**

	M	*SD*	*N*
Study 2a: China			
Monetary Incentive	32.0%		1086
Female	57.6%		1086
Non-Binary	0%		1086
Education	83.4%		1086
Current Student	72.7%		1086
No Internet Issues	35.5%		1086
Time to Load Images	5.11	12.85	1086
Age	23.31	5.78	1086
Subjective SES	4.99	1.65	1070
Study 2a: UK			
Monetary Incentive	34.4%		1067
Female	51.0%		1067
Non-Binary	1.1%		1067
Education	63.4%		1067
Current Student	11.2%		1067
Lives in a City	49.9%		1067
No Internet Issues	87.6%		1067
Time to Load Images	1.49	2.19	1067
Age	40.04	13.60	1067
Subjective SES	4.96	1.73	1067
Income	48219.11	32715.32	963
Study 2b: Mexico			

	M	*SD*	*N*
Monetary Incentive	32.6%		1053
Female	50.9%		1053
Non-Binary	2.5%		1053
Education	77.9%		1053
Current Student	59.9%		1053
Lives in a City	95.5%		1053
No Internet Issues	70.1%		1053
Time to Load Images	1.82	4.53	1053
Age	24.51	5.43	1053
Subjective SES	5.75	1.45	1053
Study 2b: US (Same Nominal Pay)			
Monetary Incentive	33.2%		1098
Female	59.4%		1098
Non-Binary	2.2%		1098
Education	65.5%		1098
Current Student	17.3%		1098
Lives in a City	68.4%		1098
No Internet Issues	87.0%		1098
Time to Load Images	1.72	2.14	1098
Age	37.28	13.77	1098
Subjective SES	5.04	1.88	1098
Income	71070.73	48702.77	1018
Study 2b: US (Same Subjective Pay)			

	M	*SD*	*N*
Monetary Incentive	33.6%		1122
Female	60.1%		1122
Non-Binary	2.2%		1122
Education	64.5%		1122
Current Student	19.9%		1122
Lives in a City	68.8%		1122
No Internet Issues	86.9%		1122
Time to Load Images	1.78	4.87	1122
Age	36.13	13.31	1122
Subjective SES	4.97	1.81	1122
Income	69599.24	47404.77	1048
Study 2c: South Africa			
Monetary Incentive	51.2%		649
Female	48.7%		649
Non-Binary	0.9%		649
Education	60.1%		649
Current Student	37.6%		649
Lives in a City	75.3%		649
No Internet Issues	79.4%		649
Time to Load Images	3.02	6.21	649
Age	28.29	7.45	649
Subjective SES	4.93	1.67	649
Income	26354.58	30553.08	502

	M	*SD*	*N*
Study 2c: US (Same Nominal Pay)			
Monetary Incentive	49.7%		662
Female	48.0%		662
Non-Binary	3.0%		662
Education	58.5%		662
Current Student	16.2%		662
Lives in a City	67.2%		662
No Internet Issues	85.2%		662
Time to Load Images	1.82	2.98	662
Age	36.91	12.62	662
Subjective SES	4.82	1.92	662
Income	72777.78	49359.34	621
Study 2c: US (Same Subjective Pay)			
Monetary Incentive	50.2%		662
Female	48.8%		662
Non-Binary	2.6%		662
Education	59.7%		662
Current Student	15.3%		662
Lives in a City	71.1%		662
No Internet Issues	83.1%		662
Time to Load Images	1.66	1.59	662
Age	36.27	12.52	662
Subjective SES	4.88	1.83	662

	M	*SD*	*N*
Income	70586.78	48008.63	605

Monetary Incentive is a binary variable for whether the participants were assigned to the monetary condition. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Current Student is a binary variable for whether participants reported being currently enrolled as students. Lives in a City is a binary variable for whether participants reported living in a medium-sized or large city. No Internet Issues is a binary variable for participants who reported having no problems with their Internet connection during the study. Time to Load Images is a scaled, continuous variable for self-reported time it took for each image to load. Age is a continuous variable for age, ranging from 18 to 70 (and above), the latter coded as 70 (for China, Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60). Subjective SES is a continuous variable for subjective socioeconomic status, ranging from 0 to 10. Income is a categorical variable, with "Income: Less than $10,000" coded as $5,000; "$10,000 to $19,999" coded as $15,000; "$20,000 to $29,999" coded as $25,000; "$30,000 to $39,999" coded as $35,000; "$40,000 to $49,999" coded as $45,000; "$50,000 to $59,999" coded as $55,000; "$60,000 to $69,999" coded as $65,000; "$70,000 to $79,999" coded as $75,000; "$80,000 to $89,999" coded as $85,000; "$90,000 to $99,999" coded as $95,000; "$100,000 to $149,999" coded as $125,000; "$150,000 or more" coded as $175,000; "Not Employed," "Student," and "Retired" were coded as missing values for the purpose of calculating the mean and standard deviation of income in each sample.

Appendix Table S29c: **Demographic Variables for Study 3a and Study 3b.**

	M	*SD*	*N*
Study 3a: India			
Monetary Incentive	49.7%		352
Female	23.6%		352
Non-Binary	0%		352
Education	95.7%		352
Current Student	2.6%		352
Lives in a City	79.5%		352
No Internet Issues	75.6%		352
Time to Load Images	3.28	5.46	352
Age	35.79	8.73	352
Subjective SES	6.07	1.64	352
Income	21911.31	25882.01	327
Study 3a: US			
Monetary Incentive	50.3%		382
Female	51.6%		382
Non-Binary	3.7%		382
Education	55.0%		382
Current Student	14.4%		382
Lives in a City	64.4%		382
No Internet Issues	81.2%		382
Time to Load Images	2.00	6.37	382
Age	37.76	13.99	382
Subjective SES	4.73	1.88	382

	M	*SD*	*N*
Income	66256.98	45403.09	358
Study 3b: US			
Monetary Incentive	50.5%		537
Female	49.0%		537
Non-Binary	1.3%		537
Education	66.9%		537
Current Student	8.4%		537
Lives in a City	66.7%		537
No Internet Issues	88.8%		537
Time to Load Images	1.68	1.77	537
Age	41.04	12.39	537
Subjective SES	4.96	1.84	537
Income	78174.90	48343.96	526

Monetary Incentive is a binary variable for whether the participants were assigned to the monetary (minimal pay) condition. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Current Student is a binary variable for whether participants reported being currently enrolled as students. Lives in a City is a binary variable for whether participants reported living in a medium-sized or large city. No Internet Issues is a binary variable for participants who reported having no problems with their Internet connection during the study. Time to Load Images is a scaled, continuous variable for self-reported time it took for each image to load. Age is a continuous variable for age, ranging from 18 to 70 (and above), the latter coded as 70. Subjective SES is a continuous variable for subjective socioeconomic status, ranging from 0 to 10. Income is a categorical variable, with "Income: Less than $10,000" coded as $5,000; "$10,000 to $19,999" coded as $15,000; "$20,000 to $29,999" coded as $25,000; "$30,000 to $39,999" coded as $35,000; "$40,000 to $49,999" coded as $45,000; "$50,000 to $59,999" coded as $55,000; "$60,000 to $69,999" coded as $65,000; "$70,000 to $79,999" coded as $75,000; "$80,000 to $89,999" coded as $85,000; "$90,000 to $99,999" coded as $95,000; "$100,000 to $149,999" coded as $125,000; "$150,000 or more" coded as $175,000; "Not Employed," "Student," and "Retired" were coded as missing values for the purpose of calculating the mean and standard deviation of income in each sample.

Appendix Table S29d: **Demographic Variables for Study 4.**

	M	*SD*	*N*
Study 4			
Monetary Incentive	50.1%		2065
English	49.6%		2065
Female	14.2%		2013
Non-Binary	0.1%		2013
Education	81.2%		2009
Current Student	77.4%		2013
Lives in a City	65.8%		2017
No Internet Issues	59.6%		2020
Time to Load Images	4.09	8.37	2019
Age	24.87	4.90	2013
Subjective SES	5.99	2.15	2009
Income (INR)	338768.12	320456.21	1104

Monetary Incentive is a binary variable for whether participants were in one monetary condition. English is a binary variable for whether the participants were assigned to complete the study in English (as compared to in Hindi). Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing an undergraduate degree or higher at the time of the experiment. Current Student is a binary variable for whether participants reported being enrolled as students at the time of the experiment. Lives in a City is a binary variable for whether participants reported living in a medium-sized or large city. No Internet Issues is a binary variable for participants who reported having no problems with their Internet connection during the study. Time to Load Images is a scaled, continuous variable for self-reported time it took for each image to load. Age is a continuous variable for age, ranging from 18 to 70 (and above), the latter coded as 70. Subjective SES is a continuous variable for subjective socioeconomic status, ranging from 0 to 10. Income is a categorical variable, with "Income: Less than ₹2.5 lakhs" coded as ₹125,000; "₹2.5 lakhs to ₹4.9 lakhs" coded as ₹375,000; "₹5 lakhs to ₹9.9 lakhs" coded as ₹750,000; "₹10 lakhs to ₹14.9 lakhs" coded as ₹1,250,000; "₹15 lakhs and above" coded as ₹1,750,000; "Not working," "Student," and "Retired" were coded as missing values for the purpose of calculating the mean and standard deviation of income in each sample.

Appendix Table S29e: Demographic Variables for Studies 5a and 5b.

Explicitly No Pay Condition (Study 5A)

Country	N	Age (SD)	Male	Current Student	Tertiary Educ	Currently Employed	Income (SD)	Median Income	SES (SD)	Urban	Ethnicity/ Race
East Africa											
Ethiopia	386	26.9 (6.9)	84.7%	52.8%	85.8%	29.0%	1.6 (1.4)	< $5,000	5.2 (1.7)	86.5%	African (76.9%)
Kenya	201	27.6 (6.1)	61.2%	32.3%	74.6%	21.4%	2.3 (2.4)	< $5,000	4.8 (1.6)	60.7%	African (91.0%)
Madagascar	175	31.2 (9.0)	50.3%	26.9%	68.0%	48.0%	1.4 (1.6)	< $5,000	5.4 (1.4)	66.9%	African (48.0%)
Uganda	177	28.8 (6.2)	58.8%	28.2%	65.0%	38.4%	1.8 (1.7)	< $5,000	4.4 (1.9)	53.1%	African (74.6%)
Zambia	173	30.4 (7.4)	60.7%	29.5%	63.0%	34.7%	1.8 (1.8)	< $5,000	4.8 (2.0)	61.3%	African (82.1%)
North Africa											
Algeria	212	30.1 (9.2)	80.7%	34.0%	86.3%	36.3%	1.8 (1.5)	< $5,000	5.4 (1.6)	68.4%	Arab (55.2%)
Egypt	238	27.8 (8.3)	65.5%	36.6%	87.0%	31.5%	1.7 (1.6)	< $5,000	5.4 (1.8)	65.1%	Arab (77.3%)
Morocco	250	29.1 (7.2)	62.4%	29.2%	75.6%	33.2%	2.0 (1.6)	$7,500	4.8 (1.8)	82.0%	Arab (58.4%)
West Africa											
Ghana	167	26.7 (6.1)	58.7%	41.9%	76.6%	44.3%	2.1 (2.1)	< $5,000	4.7 (1.6)	74.9%	African (89.8%)
Nigeria	175	30.4 (6.9)	52.6%	25.1%	93.1%	42.9%	2.5 (2.8)	< $5,000	5.1 (1.8)	70.3%	African (88.0%)
North America											
Canada	175	34.0 (11.0)	45.1%	21.1%	66.3%	74.9%	9.3 (3.1)	$75,000	5.2 (1.8)	85.1%	White (63.4%)
Jamaica	138	29.5 (9.1)	26.1%	23.2%	34.8%	39.9%	3.1 (2.4)	$7,500	4.9 (1.6)	59.4%	African (80.4%)
Mexico	211	29.2 (8.8)	64.9%	36.0%	76.8%	57.3%	4.0 (2.7)	$12,500	5.5 (1.7)	83.9%	Latino (87.7%)
US	605	38.4 (11.5)	43.5%	12.1%	66.9%	73.2%	8.3 (3.3)	$55,000	4.8 (1.9)	69.4%	White (44.8%)
South America											
Brazil	243	32.2 (9.6)	37.9%	36.6%	67.5%	60.9%	2.8 (2.5)	$7,500	5.2 (1.7)	77.4%	White (38.7%)
Chile	133	28.5 (7.4)	66.2%	51.9%	72.9%	42.1%	4.0 (3.0)	$12,500	5.3 (1.8)	87.2%	Latino (78.2%)
Colombia	190	31.3 (10.0)	57.4%	37.4%	74.7%	42.1%	2.7 (2.2)	$7,500	5.1 (1.6)	88.4%	Latino (84.2%)
Venezuela	247	38.3 (12.2)	56.7%	23.9%	73.3%	41.7%	1.8 (1.8)	< $5,000	5.1 (1.7)	81.4%	Latino (75.3%)
East Asia											
China	151	25.9 (7.6)	29.1%	57.6%	92.7%	37.7%	3.7 (2.7)	$12,500	4.7 (1.5)	68.9%	Han (95.4%)
Japan	162	43.3 (10.1)	55.6%	3.1%	66.0%	66.0%	6.1 (3.0)	$35,000	4.1 (1.9)	71.0%	Yamato (86.4%)
Central Asia											
Uzbekistan	163	26.2 (9.5)	43.6%	53.4%	70.6%	38.7%	2.7 (2.1)	$7,500	6.1 (2.1)	73.0%	Other (25.2%)
South Asia											
India	222	32.2 (9.4)	77.0%	23.9%	84.7%	52.3%	2.6 (2.3)	$7,500	6.3 (2.0)	65.3%	S. Asian (90.1%)
Sri Lanka	269	27.4 (8.7)	72.1%	58.4%	46.8%	26.0%	1.4 (1.2)	< $5,000	5.0 (1.8)	37.9%	S. Asian (84.0%)
Southeast Asia											
Indonesia	180	31.7 (7.9)	53.9%	13.9%	73.9%	50.0%	1.9 (1.5)	< $5,000	5.0 (1.6)	62.8%	S.E. Asian (90.0%)
Philippines	190	33.0 (8.9)	45.3%	10.0%	68.9%	44.7%	2.4 (2.1)	$7,500	4.9 (1.5)	58.9%	S.E. Asian (87.9%)
Vietnam	175	30.2 (8.6)	66.3%	24.0%	85.1%	53.1%	2.3 (1.6)	$7,500	5.4 (1.6)	84.0%	S.E. Asian (96.0%)
Middle East											
Jordan	127	29.5 (8.1)	57.5%	16.5%	74.8%	33.1%	2.1 (1.3)	$7,500	5.9 (1.7)	75.6%	Arab (81.1%)
Turkey	105	29.8 (9.5)	63.8%	37.1%	80.0%	41.9%	2.7 (2.1)	$7,500	5.9 (1.8)	87.6%	White (71.4%)
Eastern Europe											
Belarus	162	35.8 (11.1)	45.7%	11.1%	70.4%	58.0%	2.1 (2.0)	$7,500	5.1 (1.4)	75.9%	White (95.7%)
Hungary	115	29.7 (9.5)	64.3%	34.8%	64.3%	67.0%	4.2 (2.0)	$17,500	5.9 (1.5)	67.8%	White (97.4%)
Poland	168	27.5 (7.1)	63.1%	44.0%	72.6%	71.4%	4.8 (2.5)	$25,000	5.1 (1.6)	66.1%	White (94.6%)
Ukraine	214	38.9 (10.5)	43.5%	8.4%	74.8%	49.1%	1.8 (1.0)	< $5,000	4.7 (1.7)	76.2%	White (97.2%)
Southern Europe											
Greece	133	29.6 (8.9)	57.9%	39.8%	78.9%	52.6%	3.8 (2.2)	$12,500	5.1 (1.5)	80.5%	White (97.7%)
Italy	199	32.5 (10.0)	55.3%	38.2%	63.8%	55.3%	5.2 (2.5)	$25,000	5.4 (1.4)	53.3%	White (96.5%)
Portugal	168	28.5 (8.1)	51.2%	32.7%	82.7%	69.0%	4.7 (2.1)	$25,000	5.2 (1.3)	61.9%	White (92.9%)
Western Europe											
France	130	29.5 (9.5)	50.8%	34.6%	78.5%	60.8%	5.4 (2.8)	$25,000	5.1 (1.7)	67.7%	White (82.3%)

Germany	181	30.1 (10.2)	70.2%	42.0%	59.7%	69.1%	6.4 (3.8)	$35,000	5.4 (1.9)	50.8%	White (91.2%)
Netherlands	133	27.9 (7.4)	60.2%	38.3%	71.4%	78.2%	7.0 (3.1)	$45,000	5.8 (1.5)	63.2%	White (88.0%)
UK	224	40.7 (13.0)	43.8%	11.6%	62.1%	79.0%	7.2 (2.7)	$45,000	5.0 (1.7)	39.7%	White (86.6%)
Oceania											
Australia	156	34.3 (10.0)	51.3%	21.2%	64.1%	82.7%	9.5 (2.8)	$75,000	5.5 (1.5)	85.9%	White (75.6%)

Ambiguous Pay Condition (Study 5B)

Country	N	Age (SD)	Male	Current Student	Tertiary Educ	Currently Employed	Income (SD)	Median Income	SES (SD)	Urban	Ethnicity/ Race
East Africa											
Ethiopia	406	27.2 (6.4)	87.2%	51.5%	85.0%	32.3%	1.4 (1.1)	< $5,000	5.1 (1.7)	89.2%	African (76.6%)
Kenya	182	27.3 (5.9)	62.1%	20.9%	76.4%	23.6%	2.4 (2.8)	< $5,000	5.0 (1.4)	61.5%	African (90.1%)
Madagascar	166	30.7 (8.8)	48.8%	28.3%	67.5%	54.8%	1.3 (0.9)	< $5,000	5.3 (1.2)	74.1%	African (50.6%)
Uganda	159	28.6 (6.3)	56.6%	20.8%	70.4%	38.4%	1.8 (1.8)	< $5,000	4.5 (1.8)	58.5%	African (79.2%)
Zambia	142	29.4 (6.4)	66.9%	28.9%	59.2%	35.9%	1.5 (1.3)	< $5,000	4.6 (1.7)	72.5%	African (87.3%)
North Africa											
Algeria	188	28.5 (7.7)	77.1%	40.4%	83.0%	36.2%	1.6 (1.3)	< $5,000	5.6 (1.9)	67.6%	Arab (53.2%)
Egypt	235	28.1 (8.0)	64.3%	36.6%	88.9%	36.6%	1.7 (1.5)	< $5,000	5.7 (1.9)	66.8%	Arab (81.3%)
Morocco	249	28.4 (7.1)	61.8%	34.1%	77.5%	30.1%	1.9 (1.5)	$7,500	4.8 (1.8)	73.9%	Arab (59.8%)
West Africa											
Ghana	218	27.2 (6.8)	65.6%	33.0%	72.9%	42.7%	2.0 (2.0)	< $5,000	5.1 (1.7)	75.7%	African (83.5%)
Nigeria	147	29.8 (7.7)	42.9%	30.6%	89.1%	44.9%	3.0 (3.3)	< $5,000	5.2 (1.6)	83.7%	African (87.1%)
North America											
Canada	185	33.7 (10.5)	45.9%	23.2%	76.8%	73.5%	9.2 (3.1)	$75,000	5.3 (1.8)	82.2%	White (63.8%)
Jamaica	128	26.8 (6.7)	32.0%	30.5%	41.4%	40.6%	2.8 (2.5)	$7,500	4.8 (1.7)	50.0%	African (84.4%)
Mexico	208	30.3 (9.0)	64.4%	35.6%	76.9%	56.2%	3.7 (2.8)	$12,500	5.3 (1.7)	85.1%	Latino (87.0%)
US	620	39.0 (11.7)	44.4%	11.5%	64.8%	72.7%	8.5 (3.3)	$55,000	4.8 (1.8)	68.4%	White (48.7%)
South America											
Brazil	240	32.8 (9.5)	40.8%	38.3%	72.5%	57.1%	3.2 (2.8)	$7,500	5.2 (1.7)	81.2%	White (41.7%)
Chile	128	27.5 (6.8)	64.1%	46.1%	78.9%	46.1%	4.3 (3.3)	$12,500	5.6 (1.6)	84.4%	Latino (76.6%)
Colombia	177	30.2 (9.2)	55.9%	38.4%	73.4%	39.5%	2.4 (1.7)	$7,500	5.0 (1.9)	74.6%	Latino (87.0%)
Venezuela	246	37.8 (12.2)	56.1%	24.8%	68.3%	41.5%	1.7 (1.8)	< $5,000	5.1 (1.7)	81.3%	Latino (78.0%)
East Asia											
China	129	24.1 (5.3)	30.2%	65.9%	89.1%	32.6%	2.9 (2.0)	$12,500	4.6 (1.4)	57.4%	Han (97.7%)
Japan	154	43.5 (9.9)	51.3%	1.9%	65.6%	66.2%	5.3 (3.0)	$35,000	4.1 (2.0)	73.4%	Yamato (86.4%)
Central Asia											
Uzbekistan	162	24.7 (8.1)	54.3%	58.0%	69.1%	29.6%	2.7 (2.3)	$7,500	6.3 (2.1)	66.7%	Other (33.3%)
South Asia											
India	240	31.9 (9.1)	75.0%	18.3%	90.8%	51.2%	3.1 (2.9)	$7,500	6.2 (2.2)	65.0%	S. Asian (92.9%)
Sri Lanka	263	26.7 (8.3)	73.8%	55.5%	47.5%	28.5%	1.3 (0.6)	< $5,000	5.1 (1.7)	30.4%	S. Asian (84.0%)
Southeast Asia											
Indonesia	164	32.6 (8.1)	56.7%	7.9%	74.4%	53.7%	2.3 (2.2)	< $5,000	5.3 (1.6)	70.1%	S.E. Asian (90.2%)
Philippines	197	33.0 (9.1)	36.5%	12.2%	76.6%	42.1%	2.7 (2.4)	$7,500	5.0 (1.6)	65.0%	S.E. Asian (87.8%)
Vietnam	176	30.2 (8.4)	58.5%	23.9%	86.9%	61.9%	3.0 (2.2)	$7,500	5.5 (1.6)	87.5%	S.E. Asian (89.8%)
Middle East											
Jordan	125	30.6 (9.5)	65.6%	19.2%	84.8%	34.4%	2.2 (1.6)	$7,500	6.0 (1.6)	82.4%	Arab (76.0%)
Turkey	110	29.8 (10.4)	55.5%	37.3%	79.1%	40.0%	2.6 (2.0)	$7,500	5.9 (2.0)	90.0%	White (63.6%)
Eastern Europe											
Belarus	117	36.4 (10.5)	57.3%	11.1%	76.9%	59.0%	2.1 (1.8)	$7,500	5.2 (1.5)	65.8%	White (95.7%)
Hungary	116	29.0 (8.7)	57.8%	39.7%	76.7%	65.5%	4.3 (2.2)	$17,500	5.8 (1.4)	71.6%	White (98.3%)
Poland	153	27.6 (8.4)	62.1%	43.8%	68.6%	65.4%	4.0 (2.2)	$25,000	5.1 (1.6)	66.0%	White (98.0%)
Ukraine	209	37.6 (10.5)	46.4%	10.0%	73.7%	38.8%	2.0 (1.5)	< $5,000	4.4 (1.6)	77.5%	White (98.6%)
Southern Europe											
Greece	113	29.6 (7.8)	57.5%	40.7%	89.4%	64.6%	3.5 (2.0)	$12,500	5.3 (1.5)	83.2%	White (95.6%)

Italy	217	32.0 (10.4)	49.3%	43.8%	64.5%	58.5%	5.0 (2.4)	$25,000	5.1 (1.6)	51.2%	White (96.3%)
Portugal	170	28.3 (8.8)	53.5%	42.9%	85.9%	61.2%	4.5 (2.1)	$25,000	5.4 (1.4)	60.6%	White (95.9%)
Western Europe											
France	124	28.8 (8.7)	52.4%	29.0%	81.5%	58.1%	5.4 (2.9)	$25,000	5.3 (1.9)	62.9%	White (80.6%)
Germany	162	29.9 (9.1)	69.1%	42.0%	63.6%	72.2%	6.4 (3.2)	$35,000	5.4 (1.9)	56.8%	White (87.7%)
Netherlands	128	29.0 (8.8)	66.4%	35.9%	76.6%	77.3%	7.4 (3.3)	$45,000	6.1 (1.7)	61.7%	White (86.7%)
UK	258	41.5 (12.1)	36.4%	10.1%	60.5%	78.3%	7.2 (2.7)	$45,000	4.8 (1.7)	41.1%	White (91.1%)
Oceania											
Australia	151	31.1 (9.8)	47.0%	31.8%	68.2%	76.2%	9.5 (3.0)	$75,000	5.5 (1.6)	82.1%	White (76.8%)

Combined Sample (Studies 5A and 5B)

Country	N	Age (SD)	Male	Current Student	Tertiary Educ	Currently Employed	Income (SD)	Median Income	SES (SD)	Urban	Ethnicity/ Race
East Africa											
Ethiopia	792	27.0 (6.7)	86.0%	52.1%	85.4%	30.7%	1.5 (1.2)	< $5,000	5.1 (1.7)	87.9%	African (76.8%)
Kenya	383	27.5 (6.0)	61.6%	26.9%	75.5%	22.5%	2.3 (2.6)	< $5,000	4.9 (1.5)	61.1%	African (90.6%)
Madagascar	341	30.9 (8.9)	49.6%	27.6%	67.7%	51.3%	1.3 (1.3)	< $5,000	5.3 (1.3)	70.4%	African (49.3%)
Uganda	336	28.7 (6.3)	57.7%	24.7%	67.6%	38.4%	1.8 (1.7)	< $5,000	4.4 (1.8)	55.7%	African (76.8%)
Zambia	315	29.9 (7.0)	63.5%	29.2%	61.3%	35.2%	1.7 (1.6)	< $5,000	4.7 (1.9)	66.3%	African (84.4%)
North Africa											
Algeria	400	29.3 (8.6)	79.0%	37.0%	84.8%	36.2%	1.7 (1.4)	< $5,000	5.5 (1.7)	68.0%	Arab (54.2%)
Egypt	473	28.0 (8.1)	64.9%	36.6%	87.9%	34.0%	1.7 (1.6)	< $5,000	5.6 (1.8)	66.0%	Arab (79.3%)
Morocco	499	28.8 (7.2)	62.1%	31.7%	76.6%	31.7%	2.0 (1.6)	$7,500	4.8 (1.8)	78.0%	Arab (59.1%)
West Africa											
Ghana	385	27.0 (6.5)	62.6%	36.9%	74.5%	43.4%	2.1 (2.0)	< $5,000	4.9 (1.7)	75.3%	African (86.2%)
Nigeria	322	30.1 (7.3)	48.1%	27.6%	91.3%	43.8%	2.7 (3.0)	< $5,000	5.2 (1.7)	76.4%	African (87.6%)
North America											
Canada	360	33.8 (10.8)	45.6%	22.2%	71.7%	74.2%	9.3 (3.1)	$75,000	5.3 (1.8)	83.6%	White (63.6%)
Jamaica	266	28.2 (8.1)	28.9%	26.7%	38.0%	40.2%	3.0 (2.4)	$7,500	4.9 (1.7)	54.9%	African (82.3%)
Mexico	419	29.7 (8.9)	64.7%	35.8%	76.8%	56.8%	3.8 (2.8)	$12,500	5.4 (1.7)	84.5%	Latino (87.4%)
US	1,225	38.7 (11.6)	43.9%	11.8%	65.9%	73.0%	8.4 (3.3)	$55,000	4.8 (1.8)	68.9%	White (46.8%)
South America											
Brazil	483	32.5 (9.6)	39.3%	37.5%	70.0%	59.0%	3.0 (2.6)	$7,500	5.2 (1.7)	79.3%	White (40.2%)
Chile	261	28.0 (7.1)	65.1%	49.0%	75.9%	44.1%	4.1 (3.2)	$12,500	5.4 (1.7)	85.8%	Latino (77.4%)
Colombia	367	30.8 (9.6)	56.7%	37.9%	74.1%	40.9%	2.6 (1.9)	$7,500	5.1 (1.8)	81.7%	Latino (85.6%)
Venezuela	493	38.0 (12.2)	56.4%	24.3%	70.8%	41.6%	1.7 (1.8)	< $5,000	5.1 (1.7)	81.3%	Latino (76.7%)
East Asia											
China	280	25.1 (6.7)	29.6%	61.4%	91.1%	35.4%	3.4 (2.4)	$12,500	4.7 (1.5)	63.6%	Han (96.4%)
Japan	316	43.4 (10.0)	53.5%	2.5%	65.8%	66.1%	5.7 (3.0)	$35,000	4.1 (1.9)	72.2%	Yamato (86.4%)
Central Asia											
Uzbekistan	325	25.5 (8.8)	48.9%	55.7%	69.8%	34.2%	2.7 (2.2)	$7,500	6.2 (2.1)	69.8%	Other (29.2%)
South Asia											
India	462	32.0 (9.2)	76.0%	21.0%	87.9%	51.7%	2.8 (2.6)	$7,500	6.2 (2.1)	65.2%	S. Asian (91.6%)
Sri Lanka	532	27.0 (8.5)	72.9%	57.0%	47.2%	27.3%	1.4 (1.0)	< $5,000	5.0 (1.8)	34.2%	S. Asian (84.0%)
Southeast Asia											
Indonesia	344	32.2 (8.0)	55.2%	11.0%	74.1%	51.7%	2.1 (1.9)	< $5,000	5.1 (1.6)	66.3%	S.E. Asian (90.1%)
Philippines	387	33.0 (9.0)	40.8%	11.1%	72.9%	43.4%	2.6 (2.3)	$7,500	5.0 (1.6)	62.0%	S.E. Asian (87.9%)
Vietnam	351	30.2 (8.5)	62.4%	23.9%	86.0%	57.5%	2.7 (2.0)	$7,500	5.4 (1.6)	85.8%	S.E. Asian (92.9%)
Middle East											
Jordan	252	30.1 (8.9)	61.5%	17.9%	79.8%	33.7%	2.1 (1.5)	$7,500	5.9 (1.7)	79.0%	Arab (78.6%)
Turkey	215	29.8 (10.0)	59.5%	37.2%	79.5%	40.9%	2.6 (2.1)	$7,500	5.9 (1.9)	88.8%	White (67.4%)
Eastern Europe											
Belarus	279	36.0 (10.8)	50.5%	11.1%	73.1%	58.4%	2.1 (1.9)	$7,500	5.2 (1.4)	71.7%	White (95.7%)
Hungary	231	29.3 (9.1)	61.0%	37.2%	70.6%	66.2%	4.3 (2.1)	$17,500	5.8 (1.4)	69.7%	White (97.8%)

Poland	321	27.5 (7.7)	62.6%	43.9%	70.7%	68.5%	4.4 (2.4)	$25,000	5.1 (1.6)	66.0%	White (96.3%)
Ukraine	423	38.2 (10.5)	44.9%	9.2%	74.2%	44.0%	1.8 (1.3)	< $5,000	4.6 (1.7)	76.8%	White (97.9%)
Southern Europe											
Greece	246	29.6 (8.4)	57.7%	40.2%	83.7%	58.1%	3.7 (2.1)	$12,500	5.2 (1.5)	81.7%	White (96.7%)
Italy	416	32.2 (10.2)	52.2%	41.1%	64.2%	57.0%	5.1 (2.4)	$25,000	5.2 (1.5)	52.2%	White (96.4%)
Portugal	338	28.4 (8.5)	52.4%	37.9%	84.3%	65.1%	4.6 (2.1)	$25,000	5.3 (1.4)	61.2%	White (94.4%)
Western Europe											
France	254	29.2 (9.1)	51.6%	31.9%	79.9%	59.4%	5.4 (2.9)	$25,000	5.2 (1.8)	65.4%	White (81.5%)
Germany	343	30.0 (9.7)	69.7%	42.0%	61.5%	70.6%	6.4 (3.6)	$35,000	5.4 (1.9)	53.6%	White (89.5%)
Netherlands	261	28.5 (8.1)	63.2%	37.2%	73.9%	77.8%	7.2 (3.2)	$45,000	5.9 (1.6)	62.5%	White (87.4%)
UK	482	41.1 (12.5)	39.8%	10.8%	61.2%	78.6%	7.2 (2.7)	$45,000	4.9 (1.7)	40.5%	White (89.0%)
Oceania											
Australia	307	32.8 (10.0)	49.2%	26.4%	66.1%	79.5%	9.5 (2.9)	$75,000	5.5 (1.6)	84.0%	White (76.2%)

N is a continuous variable for the number of people in a condition or across conditions for the “Combined” table. Age is a continuous variable for age, ranging from 18 to 60 and above, the latter coded as 60. Male is a categorical variable for gender. Currently Employed and Current Student are categorical variables for whether workers reported being currently employed or enrolled as students, respectively. Tertiary Educ is a binary variable for whether workers have obtained or are in the process of obtaining a university (college) degree. Income is a categorical variable for reported annual income (in USD) with the following answer choices: 1 = “Less than $5,000;” 2 = “$5,000 to $9,999;” 3 = "$10,000 to $14,999;" 4 = "$15,000 to $19,999;" 5 = "$20,000 to $29,999;" 6 = "$30,000 to $39,999;" 7 = "$40,000 to $49,999;" 8 = "$50,000 to $59,999;" 9 = "$60,000 to $69,999;" 10 = "$70,000 to $79,999;" 11 = "$80,000 to $99,999;" 12 = "$100,000 to $149,999;" and 13 = "$150,000 or more." Median Income reflects the most commonly selected income category in each country, with the midpoint of the range presented in the table. SES is a continuous (“ladder”) variable for subjective socioeconomic status, ranging from 0 to 10. Urban is a binary variable for whether participants reported living in a medium-sized or large city. Ethnicity/Race is a variable for the most commonly reported ethnicity or race in a given sample, with simplifications such as "Black or African American" shortened to "African;" percentage of people selecting that ethnicity/race is given in parentheses.

Appendix Table S29f: Sample Characteristics for Studies 5a and 5b.

Country	N	Working on Laptop	Removed Pay Wrong	Removed Attention	Toloka	BeSample	Prolific	MTurk	Microworkers	Aishiyan	Lancers
East Africa											
Ethiopia	792	54.9%	4.2%	10.6%	95.6%	3.4%	0.0%	0.0%	1.0%	0.0%	0.0%
Kenya	383	58.5%	4.1%	7.4%	91.1%	8.9%	0.0%	0.0%	0.0%	0.0%	0.0%
Madagascar	341	70.4%	12.6%	16.7%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Uganda	336	37.8%	2.3%	10.3%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Zambia	315	30.8%	2.2%	9.8%	93.7%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%
North Africa											
Algeria	400	48.8%	9.1%	12.8%	98.8%	0.0%	0.0%	0.0%	1.2%	0.0%	0.0%
Egypt	473	37.2%	11.9%	17.2%	42.1%	56.7%	0.0%	0.0%	1.3%	0.0%	0.0%
Morocco	499	63.5%	7.2%	12.1%	89.6%	10.4%	0.0%	0.0%	0.0%	0.0%	0.0%
West Africa											
Ghana	385	40.3%	4.6%	9.6%	92.7%	5.2%	0.0%	0.0%	2.1%	0.0%	0.0%
Nigeria	322	22.4%	2.1%	6.7%	90.4%	0.0%	0.0%	0.0%	9.6%	0.0%	0.0%
North America											
Canada	360	82.2%	0.3%	0.5%	0.0%	0.0%	96.4%	3.1%	0.6%	0.0%	0.0%
Jamaica	266	40.6%	1.2%	10.0%	94.4%	0.0%	0.0%	0.4%	5.3%	0.0%	0.0%
Mexico	419	60.9%	4.0%	6.9%	49.9%	18.9%	22.7%	0.0%	8.6%	0.0%	0.0%
US	1,225	85.6%	3.7%	0.5%	0.0%	0.0%	53.1%	46.5%	0.3%	0.0%	0.0%
South America											
Brazil	483	53.6%	2.9%	5.9%	53.0%	37.5%	0.0%	4.6%	5.0%	0.0%	0.0%
Chile	261	76.6%	5.2%	10.1%	0.0%	31.0%	65.1%	0.0%	3.8%	0.0%	0.0%
Colombia	367	59.7%	8.3%	10.1%	61.3%	28.6%	0.0%	0.0%	10.1%	0.0%	0.0%
Venezuela	493	67.1%	4.6%	9.5%	79.7%	17.6%	0.0%	0.0%	2.6%	0.0%	0.0%
East Asia											
China	280	28.6%	4.8%	4.1%	0.0%	0.0%	0.0%	0.0%	6.1%	93.9%	0.0%
Japan	316	93.4%	0.3%	1.5%	0.0%	0.0%	0.0%	0.0%	0.9%	0.0%	99.1%
Central Asia											
Uzbekistan	325	34.5%	19.1%	35.0%	87.1%	12.6%	0.0%	0.0%	0.3%	0.0%	0.0%
South Asia											
India	462	59.3%	20.6%	15.0%	67.5%	4.3%	0.0%	27.9%	0.2%	0.0%	0.0%
Sri Lanka	532	72.0%	11.1%	26.2%	92.3%	0.0%	0.0%	0.0%	7.7%	0.0%	0.0%
Southeast Asia											
Indonesia	344	68.3%	6.0%	6.2%	95.6%	4.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Philippines	387	66.7%	5.5%	4.4%	99.5%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%
Vietnam	351	79.8%	10.4%	49.0%	78.3%	7.4%	0.0%	0.0%	14.2%	0.0%	0.0%
Middle East											
Jordan	252	59.9%	13.1%	18.6%	89.3%	0.0%	0.0%	0.0%	10.7%	0.0%	0.0%
Turkey	215	45.6%	8.0%	14.1%	0.0%	94.0%	0.0%	0.0%	6.0%	0.0%	0.0%
Eastern Europe											
Belarus	279	69.5%	6.1%	3.2%	87.8%	12.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Hungary	231	78.4%	1.2%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Poland	321	79.8%	8.0%	7.1%	0.0%	11.8%	85.4%	0.0%	2.8%	0.0%	0.0%
Ukraine	423	79.2%	1.7%	6.4%	90.8%	9.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Southern Europe											
Greece	246	77.2%	1.9%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Italy	416	66.3%	1.2%	1.2%	7.9%	0.7%	88.7%	1.9%	0.7%	0.0%	0.0%

Country	N	Working on Laptop	Removed Pay Wrong	Removed Attention	Toloka	BeSample	Prolific	MTurk	Microworkers	Aishiyan	Lancers
Portugal	338	88.2%	0.9%	0.6%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Western Europe											
France	254	70.1%	1.8%	0.0%	0.0%	0.4%	99.2%	0.4%	0.0%	0.0%	0.0%
Germany	343	61.2%	1.7%	1.2%	0.0%	0.0%	98.8%	1.2%	0.0%	0.0%	0.0%
Netherlands	261	72.4%	0.0%	0.7%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
UK	482	63.9%	1.0%	0.4%	0.0%	0.0%	99.8%	0.2%	0.0%	0.0%	0.0%
Oceania											
Australia	307	81.4%	1.5%	0.3%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%

N is a continuous variable for the number of people across conditions after exclusions. Working on Laptop shows the percentage of people who worked on a laptop in each country. Removed Pay Wrong and Removed Attention show the percentage of people excluded for stating that the instructions said they would receive pay or for not passing the attention check that asked them to select “2” on a scale. Toloka, BeSample, Prolific, MTurk, Microworkers, Aishiyan, and Lancers are variables that show the percentage of workers, after exclusions, in each country recruited on each of these seven platforms.

BECKER-DEGROOT-MARSHAK (BDM) PROCEDURE FOR ESTABLISHING PAY EQUIVALENCE

Study 2B

A separate group of participants in the pre-test (N = 222) completed 2 image-ratings to get a sense of the task. Then, they saw a series of choices that asked them whether they would complete 10 minutes of this task for a specified amount of money ranging from $0.10 to $6.00. They considered each pay amount separately and were told that one of their choices would be randomly implemented.

If they said "yes" to the randomly chosen amount of money, they would have to complete the whole task; if they said "no," they would not have to continue with the task but also not get extra payment. For example, if a participant said "yes" to the 80-cent question and that was the money amount that was then randomly chosen, that participant would have to complete the task and get the extra 80 cents. If they said no, they would proceed to the end of the survey and not get the extra 80 cents.

The method resulted in $0.99 expected pay for Prolific workers from the US (SD = 0.73) and $0.82 for Prolific workers in Mexico (SD = 0.87). We used these two numbers to calculate an approximate ratio of 5:6 between the two countries. Said differently, to allow for approximately the same subjective pay in the US, we would have to increase the pay in the US by 20% as compared to that in Mexico. After allowing for additional time needed for study apart from the picture task itself (such as consent and demographics), we decided to pay $1.30 in Mexico and the "Same Nominal Pay" sample in the US and $1.59 in the "Same Subjective (Increased Pay)" sample in the US. We also adjusted the additional per-piece payments in the

monetarily incentivized condition. In the increased pay sample in the US, people received 6 cents per every 10 images rated (versus 5 cents in the other two samples).

Study 2C

A total of 209 participants on Prolific from the US (N = 116) and South Africa (N = 93) passed the attention checks and proceeded to the questions that asked them how much compensation they would need to complete a 10-minute image-rating task. The procedure was identical to that in Study 2b.

Out of people who passed the attention check, 2 people in the US and 9 people in South Africa provided nonlinear responses (e.g., said they would work for $1.00 but not for $2.00). We excluded them from the equivalence calculations. The mean "requested" compensation was $1.63 in the US ($SD$ = 0.96) and $0.89 ($SD$ = 0.67) in South Africa. Thus, the ratio of anticipated pay for these samples was approximately 9:5. In South Africa and in the nominal pay sample in the United States, we retained the same pay as in the previous studies: $1.30 base pay for participation, and a 5-cent bonus per rating 10 images in the monetary condition. In the sample in the US with the same subjective pay as in South Africa, we increased the pay to $2.25, with a 9-cent bonus in the monetary condition.

Note on Purchasing Power Parity

While we relied on a procedure that has been extensively used for this purpose in previous research (Becker et al., 1964), there are multiple ways to calculate the subjective equivalence of pay between different countries. One can argue that the difference in pay between Mexico and the US sample with increased pay was too small. For instance, if we were to instead rely on the Organization for Economic and Cooperation Development 2023 conversion rates for

PPP between US and Mexico to establish subjective equivalence, we would have paid participants in the US twice as much as participants in Mexico (OECD, 2022).

However, given the fact that neither a 20% increase in base pay (Study 2b) nor an 80% increase (Study 2c) had a statistically significant effect on effort provision in the conditions where no monetary incentive was provided, we find it unlikely that raising solely the base pay will increase the amount of effort participants in the US provide in the non-monetary conditions. In fact, the interaction coefficients in both Study 2b and Study 2c are larger for the comparison with the increased (versus same nominal) pay sample, and thus it is probable that choosing an even steeper conversion rate would have only further amplified the cultural differences that we observe.

TASK COMPLETION SPEED PRE-TEST FOR STUDY 5A

A separate group of participants from Egypt, Poland, Turkey, the UK, and the US (N = 616 independent participants) completed an incentivized version of the task. Our goal was to measure the amount of time it takes people in different countries to complete the task and the proportion of people who can complete it in less than 5 minutes when they are incentivized to do it as quickly as possible. As a secondary measure, we were interested in accuracy (number of cells transcribed correctly).

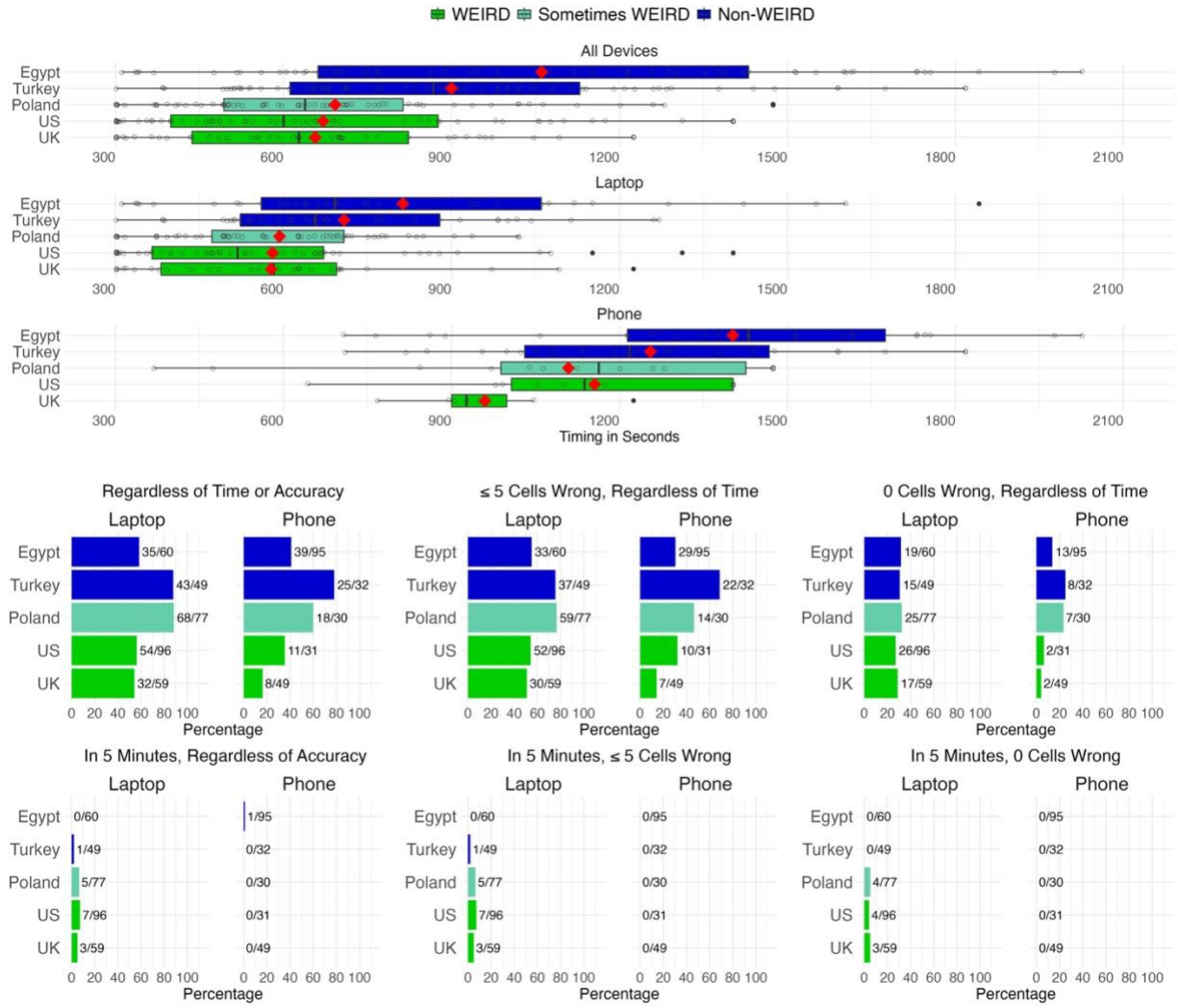
Participants were told that we were testing how quickly people could perform data entry and that they could win a bonus if they did well. The instructions will read: "Try to complete the task as quickly as possible. If you complete the task faster than most workers (above average) and with no more than one error, we will give you a $0.25 bonus. If you complete the task with no more than one error in less than 5 minutes, you will receive a $3 bonus." Thus, the workers had an incentive to complete the task within the 5-minute limit but also to continue past it, if they were unable to complete it in 5 minutes.

The top panel of Fig. S3 shows that workers in the UK, US, and Poland were able to complete the whole task rather accurately (making five or fewer errors in the transcription) faster than workers in Turkey and particularly in Egypt, $Mean_{UK}$ = 669.3 seconds, 95% CI 582.2 to 756.4; $Mean_{US}$ = 693.9, 95% CI 594.6 to 791.3; $Mean_{Poland}$ = 699.8, 95% CI 626.5 to 773.12; $Mean_{Turkey}$ = 905.1, 95% CI 813.0 to 997.1; $Mean_{Egypt}$ = 1248.9, 95% CI 1061.3 to 1436.5. It took a worker in Turkey, on average, more than 3 minutes longer than a worker in one of the Western countries we sampled; it took a worker in Egypt nearly twice as long.

The bottom panel of Fig. S3 shows that few people could complete the transcription in less than five minutes and with five or fewer cells inputted incorrectly (0 out of 155 participants

in Egypt across laptop and phone, 1 out of 81 in Turkey, 5 out of 107 in Poland, 7 out of 127 in the US, and 3 out of 108 in the UK). No participant using a phone was able to complete the whole task within the five-minute limit with five or fewer cells inputted incorrectly.

Fig. S3: **Speed and Likelihood of Completing the Whole Task When Incentivized, by Completion Time (In 5 Minutes or Not) and Accuracy.**



The top panel shows the number of seconds it takes participants in five countries—Egypt (*N* = 155 participants recruited on Toloka), Turkey (*N* = 81 participants recruited on BeSample), Poland (*N* = 107 participants recruited on Prolific), the US (*N* = 127 participants recruited on Prolific), and UK (*N* = 108 participants recruited on Prolific)—to complete the task inputting five or fewer cells incorrectly, across all devices, on their laptops, or phones. Participants who completed the task on devices other than laptop or phone (*N* = 38) are excluded beyond the "All Devices" part of the top panel and only participants who completed with five errors or fewer are shown. The black line within each box represents a median, and the red rhombus shows a mean. For mean speed, we winsorized the individual times of participants at 2.5 SD above/below the mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The bottom panel represents the proportion of participants completing the task grouped by device type (laptop or phone), 4 (in any amount of time or within the five-minute limit), and accuracy (with any number of cells inputted incorrectly, with 5 cells or fewer inputted incorrectly, and with 0 cells inputted incorrectly).

ADDITIONAL CONTROLS IN STUDIES 2A–4

Repeating the Analyses with Interactions Between Country and Each Predictor Variable

Following the methods proposed by Bandiera and colleagues (Bandiera et al., 2020) and suggestions from an anonymous reviewer, we first repeated the analyses in the main text with interactions between country and each of the principal demographic predictor variables for all the main analyses that involved comparisons across samples. The results are presented in Appendix Tables S30a and S30b.

Appendix Table S30a: **Studies 2a–c, Mean Effort in Monetary Versus Pooled Non-Monetary Incentive Conditions, with Interactions Between Country and Predictor Variables.**

	b	*SE*	*t*	*P*	95% CI
Study 2a: China and the UK					
Monetary Incentive	8.88	2.83	3.14	0.002	[3.34, 14.42]
UK	5.09	4.69	1.09	0.278	[-4.11, 14.29]
Female	5.03	2.68	1.88	0.060	[-0.21, 10.28]
Non-Binary	8.76	12.69	0.69	0.490	[-16.12, 33.65]
Education	-3.55	3.93	-0.91	0.365	[-11.25, 4.14]
Age	1.97	3.36	0.59	0.558	[-4.62, 8.57]
Monetary Incentive × UK	38.36	3.98	9.64	< 0.001	[30.56, 46.16]
Female × UK	-7.73	3.78	-2.04	0.041	[-15.15, -0.31]
Education × UK	-1.40	4.82	-0.29	0.772	[-10.84, 8.05]
Age × UK	-5.35	3.61	-1.48	0.139	[-12.44, 1.73]
Study 2b: Mexico and the US (Same Nominal Pay)					
Monetary Incentive	27.05	2.92	9.28	< 0.001	[21.33, 32.77]
US	-41.27	4.83	-8.54	< 0.001	[-50.75, -31.80]
Female	-12.22	2.79	-4.38	< 0.001	[-17.69, -6.76]
Non-Binary	9.86	8.94	1.10	0.271	[-7.68, 27.40]
Education	-2.85	3.30	-0.86	0.388	[-9.31, 3.61]
Age	-2.75	3.13	-0.88	0.378	[-8.88, 3.38]
Monetary Incentive × US	14.45	4.07	3.55	< 0.001	[6.48, 22.43]
Female × US	14.36	3.93	3.66	< 0.001	[6.66, 22.07]
Non-Binary × US	4.92	12.93	0.38	0.704	[-20.44, 30.28]
Education × US	-5.16	4.34	-1.19	0.234	[-13.67, 3.35]
Age × US	1.53	3.35	0.46	0.648	[-5.04, 8.10]
Study 2b: Mexico and the US (Same Subjective Pay)					
Monetary Incentive	27.05	2.95	9.17	< 0.001	[21.26, 32.84]
US	-40.74	4.86	-8.38	< 0.001	[-50.28, -31.21]
Female	-12.22	2.82	-4.33	< 0.001	[-17.76, -6.69]
Non-Binary	9.86	9.05	1.09	0.276	[-7.90, 27.61]

	b	*SE*	*t*	*P*	95% CI
Education	-2.85	3.34	-0.85	0.393	[-9.39, 3.69]
Age	-2.64	3.03	-0.87	0.384	[-8.58, 3.30]
Monetary Incentive × US	20.04	4.09	4.90	< 0.001	[12.01, 28.06]
Female × US	13.30	3.98	3.34	< 0.001	[5.50, 21.11]
Non-Binary × US	3.42	12.97	0.26	0.792	[-22.01, 28.85]
Education × US	-4.55	4.37	-1.04	0.298	[-13.12, 4.02]
Age × US	-0.05	3.26	-0.02	0.988	[-6.45, 6.35]
Study 2c: South Africa and the US (Same Nominal Pay)					
Monetary Incentive	25.15	2.88	8.75	< 0.001	[19.51, 30.79]
US	-11.49	4.36	-2.63	0.009	[-20.05, -2.93]
Female	-3.89	2.92	-1.33	0.183	[-9.63, 1.84]
Non-Binary	-13.33	15.10	-0.88	0.377	[-42.95, 16.29]
Education	-4.09	2.97	-1.38	0.169	[-9.91, 1.74]
Age	1.24	2.17	0.57	0.569	[-3.03, 5.50]
Monetary Incentive × US	8.42	4.06	2.08	0.038	[0.46, 16.38]
Female × US	8.59	4.13	2.08	0.038	[0.48, 16.70]
Non-Binary × US	9.32	17.30	0.54	0.590	[-24.63, 43.27]
Education × US	-5.18	4.15	-1.25	0.212	[-13.33, 2.97]
Age × US	-5.67	2.52	-2.24	0.025	[-10.62, -0.71]
Study 2c: South Africa and the US (Same Subjective Pay)					
Monetary Incentive	25.15	2.93	8.59	< 0.001	[19.40, 30.90]
US	-13.29	4.43	-3.00	0.003	[-21.98, -4.61]
Female	-3.89	2.98	-1.31	0.191	[-9.73, 1.95]
Non-Binary	-13.33	15.38	-0.87	0.386	[-43.51, 16.84]
Education	-4.09	3.03	-1.35	0.177	[-10.02, 1.85]
Age	1.22	2.18	0.56	0.576	[-3.06, 5.49]
Monetary Incentive × US	15.69	4.13	3.80	< 0.001	[7.60, 23.79]
Female × US	3.58	4.19	0.85	0.393	[-4.64, 11.81]
Non-Binary × US	1.00	17.97	0.06	0.956	[-34.25, 36.24]

	b	*SE*	*t*	*P*	95% CI
Education × US	0.14	4.25	0.03	0.973	[-8.19, 8.47]
Age × US	-3.08	2.54	-1.21	0.225	[-8.06, 1.90]

Regression output for the linear regression models on the number of images rated in the samples from Studies 2a–c for each sample comparison pair, with interactions between country and each predictor variable added. Monetary Incentive is a binary variable for whether participants were in one of the monetary conditions. UK and US are binary variable for whether participants were part of the UK or one of the US samples, respectively. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Age is a continuous variable and is scaled.

Appendix Table S30b: **Study 3a, Mean Effort in Monetary Versus Pooled Non-Monetary Incentive Conditions, with Interactions Between Country and Predictor Variables.**

	b	*SE*	*t*	*P*	95% CI
Study 3a: India and the US					
Monetary Incentive	1.47	4.29	0.34	0.732	[-6.96, 9.90]
US	-17.00	11.44	-1.49	0.138	[-39.45, 5.45]
Female	3.54	5.04	0.70	0.483	[-6.35, 13.43]
Non-Binary	19.52	11.19	1.74	0.081	[-2.44, 41.49]
Education	3.47	10.61	0.33	0.744	[-17.36, 24.30]
Age	4.52	2.91	1.55	0.122	[-1.21, 10.24]
Monetary Incentive × US	13.12	5.96	2.20	0.028	[1.43, 24.81]
Female × US	-1.76	6.56	-0.27	0.788	[-14.65, 11.12]
Education × US	-9.21	11.39	-0.81	0.419	[-31.57, 13.16]
Age × US	-6.67	3.40	-1.96	0.050	[-13.34, 0.00]

Regression output for the linear regression models on the number of images rated in the Study 3a samples from India (N = 352) and the US (N = 382), with interactions between country and each predictor variable added. Monetary Incentive is a binary variable for whether participants were assigned to the monetary (minimal pay) condition. US is a binary variable for whether participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Age is a continuous variable and is scaled.

Repeating the Analyses in Studies 2a–c and 3a with Additional Controls

For Study 2c, we pre-registered controlling for age, gender, and education only. However, the samples in South Africa and the US differed on some other demographic dimensions, such as the proportion of then-current students and that of people living in cities (as opposed to towns and villages). Similarly, we did not pre-register controlling for quality of the Internet connection in Study 2c. However, it is possible that having a slow connection would have prevented participants from rating more images, and that our samples differed in their Internet speeds.

Thus, in all the studies with between-culture comparisons for which we collected data (Studies 2a–c and 3a), we re-ran the regressions on the mean number of images rated with additional controls. We wanted to test whether the principal interaction between culture and incentive type (monetary or non-monetary) would remain robust. None of the main results changed. We report the regression outputs in Appendix Tables 31a and S31b.

Additional Demographic Variables

Student Status and Living in a City

We added to the regression model binary variables for (1) whether participants were then-students and (2) whether they lived in a medium- or large-sized city.

Subjective SES

We also included participants' self-reported subjective SES, measured on the 0–10 "ladder" scale, where 0 = "0 (bottom) and 10 = "10 (top)." Participants saw an image of a ladder and were asked, "Think of this ladder as representing people's social status in your country. At

the top of the ladder are the people who are the best off—those who have the most money, the most education, and the most respected jobs. At the bottom are the people who are the worst off—who have the least money, the least education, and the least respected jobs or no job. Where would you place yourself on this ladder compared to people in your country?"

Income

We measured income on a scale that included non-numerical options such as "Not currently employed," "Retired," and "Student." In Studies 2a–c, 3a, and Study 3b, numerical ranges for the income question were "Income: Less than $10,000" coded as $5,000; "$10,000 to $19,999" coded as $15,000; "$20,000 to $29,999" coded as $25,000; "$30,000 to $39,999" coded as $35,000; "$40,000 to $49,999" coded as $45,000; "$50,000 to $59,999" coded as $55,000; "$60,000 to $69,999" coded as $65,000; "$70,000 to $79,999" coded as $75,000; "$80,000 to $89,999" coded as $85,000; "$90,000 to $99,999" coded as $95,000; "$100,000 to $149,999" coded as $125,000; and "$150,000 or more" coded as $175,000.

We did not measure income as well as we could have. First—in all studies except for Study 4—we employed the scale we use by default in the US, with the lowest cut-off being $10,000 per year. $10,000 is higher than the national net income per capita of all four "non-WEIRD" countries in our samples (World Bank, 2020).

Second, in China, we asked about income as a text-entry question instead, which made analysis unfeasible. Third, in Mexico, we did not specify whether "$" refers to the American dollar or the Mexican peso, which led to uninterpretable data. For these reasons, we did not include descriptive statistics for income for the samples from China and Mexico in the Demographic Variables section. For the same reasons, in Study 2a, we did not include income in

the regressions with additional controls, and in Study 2b, we included only non-numerical options ("Not currently employed," "Retired," and "Student").

Internet Connection

The first Internet variable asked participants, "Did you have any internet connection/speed issues during the study?" and was measured on a scale from 1 ("No issues at all") to 5 ("Lots of Issues"). We then created a binary variable called "No Internet Issues," where 1 = 1 ("No issues at all") and 0 = all other values of the variable.

The second variable, "Time to Load Images," asked, "On average, how many seconds did it take for the picture to load completely? You may include up to 1 decimal point." Participants were then free to input any number below 120 (2 minutes).

Appendix Table S31a: **Studies 2a–c, Mean Effort in Monetary Versus Pooled Non-Monetary Incentive Conditions, with Additional Controls.**

	b	*SE*	*t*	*P*	95% CI
Study 2a: China and the UK					
Monetary Incentive	8.37	2.85	2.94	0.003	[2.78, 13.96]
UK	1.32	3.13	0.42	0.674	[-4.83, 7.46]
Female	1.01	1.91	0.53	0.597	[-2.73, 4.75]
Non-Binary	9.87	12.67	0.78	0.436	[-14.98, 34.72]
Education	-5.36	2.36	-2.27	0.023	[-9.99, -0.73]
Current Student	-1.57	2.77	-0.57	0.571	[-7.01, 3.87]
Age	-2.61	1.32	-1.97	0.049	[-5.21, -0.02]
No Internet Issues	-4.11	2.31	-1.78	0.075	[-8.64, 0.42]
Time to Load Images	-3.60	0.96	-3.74	< 0.001	[-5.49, -1.71]
Subjective SES	-0.85	0.97	-0.88	0.378	[-2.75, 1.04]
Monetary Incentive × UK	38.75	4.00	9.69	< 0.001	[30.91, 46.59]
Study 2b: Mexico and the US (Same Nominal Pay)					
Monetary Incentive	29.25	3.14	9.32	< 0.001	[23.10, 35.41]
US	-38.98	2.92	-13.34	< 0.001	[-44.72, -33.25]
Female	-5.20	2.04	-2.55	0.011	[-9.20, -1.19]
Non-Binary	9.87	6.75	1.46	0.144	[-3.38, 23.11]
Education	-4.80	2.36	-2.03	0.042	[-9.42, -0.17]
Current Student	1.68	2.57	0.65	0.513	[-3.36, 6.72]
No Internet Issues	-5.37	2.51	-2.14	0.032	[-10.29, -0.46]
Time to Load Images	-0.94	1.03	-0.92	0.359	[-2.96, 1.07]
Lives in a City	-7.41	2.72	-2.73	0.006	[-12.74, -2.08]
Age	-1.57	1.28	-1.23	0.219	[-4.07, 0.93]
Subjective SES	-1.11	1.07	-1.03	0.302	[-3.21, 1.00]
Income: Not Employed	1.35	5.31	0.25	0.799	[-9.06, 11.75]
Income: Retired	-18.95	18.29	-1.04	0.300	[-54.82, 16.93]
Monetary Incentive × US	12.95	4.25	3.05	0.002	[4.62, 21.28]
Study 2b: Mexico and the US (Same Subjective Pay)					

	b	*SE*	*t*	*P*	95% CI
Monetary Incentive	28.96	3.18	9.12	< 0.001	[22.73, 35.20]
US	-37.35	2.92	-12.79	< 0.001	[-43.08, -31.62]
Female	-6.13	2.08	-2.95	0.003	[-10.20, -2.06]
Non-Binary	7.53	6.76	1.11	0.265	[-5.72, 20.77]
Education	-5.33	2.35	-2.27	0.023	[-9.94, -0.72]
Current Student	-1.32	2.53	-0.52	0.603	[-6.28, 3.65]
No Internet Issues	-8.07	2.53	-3.19	0.001	[-13.03, -3.11]
Time to Load Images	-1.87	1.00	-1.86	0.063	[-3.83, 0.10]
Lives in a City	-7.07	2.74	-2.58	0.010	[-12.44, -1.69]
Age	-3.07	1.26	-2.44	0.015	[-5.53, -0.60]
Subjective SES	1.43	1.08	1.32	0.187	[-0.69, 3.56]
Income: Not Employed	-3.91	5.64	-0.69	0.488	[-14.97, 7.15]
Income: Retired	-17.91	20.14	-0.89	0.374	[-57.41, 21.59]
Monetary Incentive × US	17.46	4.27	4.08	< 0.001	[9.07, 25.84]
Study 2c: South Africa and the US (Same Nominal Pay)					
Monetary Incentive	25.33	3.08	8.23	< 0.001	[19.29, 31.37]
US	-9.74	3.47	-2.81	0.005	[-16.54, -2.94]
Female	0.96	2.17	0.44	0.657	[-3.29, 5.22]
Non-Binary	-6.70	7.62	-0.88	0.380	[-21.65, 8.26]
Education	-6.76	2.34	-2.89	0.004	[-11.36, -2.16]
Current Student	-3.53	2.85	-1.24	0.216	[-9.11, 2.06]
No Internet Issues	-2.13	2.82	-0.75	0.452	[-7.66, 3.41]
Time to Load Images	-1.23	1.05	-1.17	0.243	[-3.28, 0.83]
Lives in a City	-1.92	2.40	-0.80	0.423	[-6.63, 2.78]
Age	-3.50	1.25	-2.80	0.005	[-5.95, -1.04]
Subjective SES	0.49	1.24	0.40	0.693	[-1.94, 2.92]
Income: $10,000 to $19,999	0.24	4.11	0.06	0.953	[-7.82, 8.31]
Income: $20,000 to $29,999	-1.22	4.14	-0.29	0.769	[-9.35, 6.91]
Income: $30,000 to $39,999	-0.32	4.64	-0.07	0.944	[-9.43, 8.78]

	b	*SE*	*t*	*P*	95% CI
Income: $40,000 to $49,999	-1.69	4.85	-0.35	0.727	[-11.20, 7.82]
Income: $50,000 to $59,999	-2.80	5.55	-0.51	0.614	[-13.70, 8.09]
Income: $60,000 to $69,999	0.54	5.64	0.10	0.924	[-10.52, 11.60]
Income: $70,000 to $79,999	-2.18	6.42	-0.34	0.734	[-14.77, 10.41]
Income: $80,000 to $89,999	3.65	7.55	0.48	0.629	[-11.17, 18.47]
Income: $90,000 to $99,999	5.97	6.43	0.93	0.354	[-6.66, 18.59]
Income: $100,000 to $149,999	-0.73	5.36	-0.14	0.891	[-11.25, 9.78]
Income: $150,000 or more	-7.67	6.04	-1.27	0.204	[-19.52, 4.17]
Income: Not Employed	-2.64	7.24	-0.36	0.715	[-16.85, 11.56]
Income: Student	-1.99	5.43	-0.37	0.715	[-12.65, 8.67]
Income: Retired	2.49	19.10	0.13	0.896	[-34.98, 39.96]
Monetary Incentive × US	8.44	4.25	1.98	0.047	[0.10, 16.78]
Study 2c: South Africa and the US (Same Subjective Pay)					
Monetary Incentive	24.51	3.12	7.87	< 0.001	[18.39, 30.62]
US	-9.77	3.50	-2.79	0.005	[-16.63, -2.90]
Female	-2.18	2.20	-0.99	0.322	[-6.49, 2.14]
Non-Binary	-13.31	8.19	-1.63	0.104	[-29.38, 2.75]
Education	-3.30	2.44	-1.36	0.176	[-8.09, 1.48]
Current Student	-4.69	3.01	-1.56	0.120	[-10.60, 1.22]
No Internet Issues	-7.40	2.78	-2.66	0.008	[-12.86, -1.95]
Time to Load Images	-1.78	1.06	-1.67	0.096	[-3.86, 0.31]
Lives in a City	-0.67	2.47	-0.27	0.786	[-5.53, 4.18]
Age	-1.15	1.26	-0.92	0.359	[-3.62, 1.32]
Subjective SES	0.46	1.25	0.36	0.716	[-2.00, 2.91]
Income: $10,000 to $19,999	-0.19	4.11	-0.05	0.964	[-8.25, 7.88]
Income: $20,000 to $29,999	-2.59	4.27	-0.60	0.545	[-10.97, 5.80]
Income: $30,000 to $39,999	-2.14	4.71	-0.45	0.650	[-11.38, 7.11]
Income: $40,000 to $49,999	3.14	4.92	0.64	0.524	[-6.51, 12.78]
Income: $50,000 to $59,999	-9.86	5.37	-1.84	0.067	[-20.39, 0.68]

	b	*SE*	*t*	*P*	95% CI
Income: $60,000 to $69,999	-1.68	5.67	-0.30	0.768	[-12.79, 9.44]
Income: $70,000 to $79,999	-9.80	6.35	-1.54	0.123	[-22.27, 2.67]
Income: $80,000 to $89,999	-7.71	7.31	-1.05	0.292	[-22.05, 6.63]
Income: $90,000 to $99,999	-7.92	7.29	-1.09	0.278	[-22.23, 6.39]
Income: $100,000 to $149,999	-6.69	5.30	-1.26	0.207	[-17.09, 3.71]
Income: $150,000 or more	-10.58	6.37	-1.66	0.097	[-23.08, 1.92]
Income: Not Employed	-1.72	7.67	-0.22	0.822	[-16.76, 13.32]
Income: Student	-0.61	5.28	-0.11	0.909	[-10.97, 9.76]
Income: Retired	-15.91	17.20	-0.93	0.355	[-49.66, 17.84]
Monetary Incentive × US	16.28	4.30	3.78	< 0.001	[7.84, 24.73]

Regression output for the linear regression models on the number of images rated in the samples from Studies 2a–c for each sample comparison pair, with additional controls added. Monetary Incentive is a binary variable for whether participants were in one of the monetary conditions. UK and US are binary variable for whether participants were part of the UK or one of the US samples, respectively. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. No Internet Issues is a binary variable for participants who reported having no problems with their Internet connection during the study. Time to Load Images is a scaled, continuous variable for self-reported average time to load each image. Live in Cities is a binary variable that specifies whether a participant lives in a medium- or large-sized city. Age is a continuous variable and is scaled. Subjective SES is a continuous variable for subjective socioeconomic status, ranging from 0 to 10. Income is a categorical variable, with "Income: Less than $10,000" coded as $5,000; "$10,000 to $19,999" coded as $15,000; "$20,000 to $29,999" coded as $25,000; "$30,000 to $39,999" coded as $35,000; "$40,000 to $49,999" coded as $45,000; "$50,000 to $59,999" coded as $55,000; "$60,000 to $69,999" coded as $65,000; "$70,000 to $79,999" coded as $75,000; "$80,000 to $89,999" coded as $85,000; "$90,000 to $99,999" coded as $95,000; "$100,000 to $149,999" coded as $125,000; and "$150,000 or more" coded as $175,000. Income also included non-numerical categories of "Not Employed," "Student," and "Retired."

Appendix Table S31b: **Study 3a, Mean Effort in Norm Versus Minimal Pay Condition, with Additional Controls.**

	b	*SE*	*t*	*P*	95% CI
Study 3a: India and the US					
Monetary Incentive	-0.54	4.41	-0.12	0.903	[-9.19, 8.11]
US	-27.50	5.92	-4.65	< 0.001	[-39.12, -15.88]
Female	1.42	3.29	0.43	0.667	[-5.05, 7.89]
Non-Binary	22.10	11.53	1.92	0.056	[-0.54, 44.73]
Education	-0.80	4.32	-0.18	0.853	[-9.27, 7.67]
Current Student	-3.93	6.16	-0.64	0.524	[-16.02, 8.16]
No Internet Issues	-8.67	3.76	-2.30	0.021	[-16.05, -1.28]
Time to Load Images	-1.33	1.49	-0.89	0.373	[-4.27, 1.60]
Lives in a City	-2.35	3.47	-0.68	0.499	[-9.16, 4.46]
Age	-0.57	1.67	-0.34	0.733	[-3.84, 2.71]
Subjective SES	-5.85	1.93	-3.04	0.002	[-9.63, -2.07]
Income: $10,000 to $19,999	-7.57	5.17	-1.46	0.144	[-17.73, 2.58]
Income: $20,000 to $29,999	-17.15	5.72	-3.00	0.003	[-28.39, -5.91]
Income: $30,000 to $39,999	-10.19	6.45	-1.58	0.115	[-22.86, 2.48]
Income: $40,000 to $49,999	-9.13	7.71	-1.18	0.237	[-24.27, 6.01]
Income: $50,000 to $59,999	-2.89	6.81	-0.43	0.671	[-16.27, 10.48]
Income: $60,000 to $69,999	-10.68	8.02	-1.33	0.183	[-26.43, 5.06]
Income: $70,000 to $79,999	4.34	7.95	0.55	0.585	[-11.27, 19.96]
Income: $80,000 to $89,999	-12.36	8.87	-1.39	0.164	[-29.76, 5.05]
Income: $90,000 to $99,999	-4.50	8.84	-0.51	0.611	[-21.86, 12.86]
Income: $100,000 to $149,999	4.76	8.16	0.58	0.560	[-11.26, 20.78]
Income: $150,000 or more	-4.71	9.70	-0.49	0.627	[-23.76, 14.33]
Income: Not Employed	-14.07	12.13	-1.16	0.246	[-37.89, 9.74]
Income: Student	-6.62	16.70	-0.40	0.692	[-39.41, 26.16]
Income: Retired	23.44	28.72	0.82	0.415	[-32.95, 79.82]
Monetary Incentive × US	14.61	6.10	2.39	0.017	[2.63, 26.59]

Regression output for the linear regression models on the number of images rated in Study 3a, with additional controls added. Monetary Incentive is a binary variable for whether participants were assigned to the monetary (minimal pay) condition. US is a binary variable for whether participants were part of the US sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Current Student is a binary variable for those who report being currently enrolled as students. Lives in a City is a binary variable for participants who reported living in a medium-sized or large city. No Internet Issues is a binary variable for participants who reported having no problems with their Internet connection during the study. Time to Load Images a continuous variable for self-reported time it took for each image to load. Age is a continuous variable and is scaled. Subjective SES is a continuous variable for subjective socioeconomic status, ranging from 0 to 10. Income is a categorical variable, with "Income: Less than $10,000" coded as $5,000; "$10,000 to $19,999" coded as $15,000; "$20,000 to $29,999" coded as $25,000; "$30,000 to $39,999" coded as $35,000; "$40,000 to $49,999" coded as $45,000; "$50,000 to $59,999" coded as $55,000; "$60,000 to $69,999" coded as $65,000; "$70,000 to $79,999" coded as $75,000; "$80,000 to $89,999" coded as $85,000; "$90,000 to $99,999" coded as $95,000; "$100,000 to $149,999" coded as $125,000; and "$150,000 or more" coded as $175,000. Income also included non-numerical categories of "Not Employed," "Student," and "Retired."

ADDITIONAL ANALYSES FOR STUDY 3A

As specified in the Methods section for Study 3a, we pre-registered t-tests to compare the two incentive conditions within each country, with a prediction that the two incentives would be statistically indistinguishable in India but not in the US.

In the US, participants completed more images in the minimal pay compared to the norm condition, two-sided Welch's $t(354.7) = 3.45$, $P < 0.001$, $Mean_{difference} = 14.91$, 95% CI 6.41 to 23.41. In India, there was no statistically significant difference, two-sided Welch's $t(349.02) = 0.21$, $P = 0.834$, $Mean_{difference} = 0.84$, 95% CI -7.04 to 8.71.

The pre-registered analysis also specified recruiting 352 participants per country post-exclusions. Since fewer people than we had thought did not pass the checks in the US, we ended up with a final sample of 382 participants. Below we repeat the analysis reported in the paper, including only the first 352 participants in the US.

In the US, the *t*-test for differences between the two incentive conditions on the number of images rated remains statistically significant, two-sided Welch's $t(316.74) = 3.89$, $P < 0.001$, $Mean_{difference} = 17.17$, 95% CI 8.48 to 25.87.

The interaction between culture and incentive, reported in the main test as $b = 13.77$, $t(726) = 2.31$, $P = 0.021$, 95% CI 2.09 to 25.45, becomes $b = 16.11$, $t(696) = 2.68$, $P = 0.007$, 95% CI 4.32 to 27.89.

EXPLORATORY VARIABLES IN STUDIES 2–4

In the samples that we recruited relatively early in the project (Studies 2a–b and Study 4), we included 7 exploratory variables that asked participants to report their motivation for completing the task and satisfaction with their performance, all on a 7-point Likert scale (1 = "Strongly Disagree" to 7 = "Strongly Agree"). We did not analyze these variables until after we had made the (wrong) decision to exclude them from future studies to minimize the time it takes to complete the survey. We report these variables for samples in Studies 2a, 2b, and 4 in Appendix Tables S32, S33a, S33b, and S34. Results for Study 4 are also presented in Fig. S4.

Additional Variables Across Studies

In the earlier studies we ran, we collected more items measured on the same scale as the 7 exploratory DVs. For instance, we gave participants the relational mobility scale (Thomson et al., 2018). We did this to adjust for different cultural response styles by ipsatizing (Bartram, 1996) the 7 exploratory variables. Ipsatization is a procedure used in cultural and personality psychology that adjusts people's responses to remove the influence of cultural differences in response styles. Ipsatization removes the mean of the items measured on the same scale but measuring different constructs from each variable of interest. We only analyzed the exploratory variables in the main text in Study 4, where no adjustments for different cultural response styles were needed (because participants came from the same culture—India). Thus, we did not use any of these additional measures. However, they are available in the study materials from Qualtrics.

Individual Differences in Study 4

In Study 4, we first collected data from numerous participants in India on a number of cultural scales—tightness/looseness (Gelfand et al., 2011; Jackson et al., 2020), relational

mobility (Thomson et al., 2018), self-construal (Singelis, 1994)—for a different project. A week later, some of these participants completed the task for Study 4. We considered studying how individual differences might affect how many images people rate in the monetary and norm conditions in India. However, we ultimately decided against the idea of including the discussion of individual differences and did not analyze these data for this project.

Appendix Table S32: **Exploratory Variables in Study 2a.**

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
I enjoyed completing the task						
(Intercept)	5.88	0.07	78.79	< 0.001	< 0.001	[5.68, 6.09]
Monetary Incentive	0.07	0.08	0.92	0.359	1.000	[-0.14, 0.29]
UK	-0.60	0.08	-7.68	< 0.001	< 0.001	[-0.81, -0.39]
Female	0.04	0.05	0.82	0.415	1.000	[-0.10, 0.19]
Non-Binary	-0.11	0.36	-0.31	0.753	1.000	[-1.08, 0.85]
Education	-0.31	0.06	-4.85	< 0.001	< 0.001	[-0.48, -0.14]
Age	0.17	0.03	5.00	< 0.001	< 0.001	[0.08, 0.27]
Monetary Incentive × UK	0.17	0.11	1.50	0.134	0.941	[-0.14, 0.47]
I am satisfied with how well I did on the task						
(Intercept)	5.93	0.07	90.73	< 0.001	< 0.001	[5.75, 6.10]
Monetary Incentive	-0.01	0.07	-0.13	0.900	1.000	[-0.20, 0.18]
UK	0.10	0.07	1.41	0.158	1.000	[-0.09, 0.28]
Female	-0.12	0.05	-2.57	0.010	0.071	[-0.25, 0.01]
Non-Binary	0.14	0.31	0.44	0.659	1.000	[-0.71, 0.98]
Education	-0.07	0.06	-1.28	0.201	1.000	[-0.22, 0.08]
Age	0.08	0.03	2.65	0.008	0.056	[0.00, 0.16]
Monetary Incentive × UK	0.16	0.10	1.60	0.111	0.776	[-0.11, 0.42]
I believe that I helped others by completing the task						
(Intercept)	6.12	0.07	82.09	< 0.001	< 0.001	[5.92, 6.32]
Monetary Incentive	-0.05	0.08	-0.58	0.565	1.000	[-0.26, 0.17]
UK	-1.08	0.08	-13.84	< 0.001	< 0.001	[-1.29, -0.87]
Female	-0.03	0.05	-0.53	0.594	1.000	[-0.17, 0.12]
Non-Binary	-0.34	0.36	-0.95	0.342	1.000	[-1.30, 0.63]
Education	-0.15	0.06	-2.43	0.015	0.105	[-0.32, 0.02]
Age	0.08	0.03	2.29	0.022	0.156	[-0.01, 0.17]
Monetary Incentive × UK	0.23	0.11	2.02	0.044	0.306	[-0.08, 0.53]
Completing the task was boring						
(Intercept)	2.95	0.09	31.56	< 0.001	< 0.001	[2.7, 3.2]

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
Monetary Incentive	-0.06	0.10	-0.60	0.552	1.000	[-0.33, 0.21]
UK	0.42	0.10	4.33	< 0.001	< 0.001	[0.16, 0.69]
Female	-0.12	0.07	-1.77	0.077	0.540	[-0.30, 0.06]
Non-Binary	-0.26	0.45	-0.57	0.568	1.000	[-1.47, 0.95]
Education	0.20	0.08	2.57	0.010	0.071	[-0.01, 0.42]
Age	-0.21	0.04	-4.82	< 0.001	< 0.001	[-0.33, -0.09]
Monetary Incentive × UK	0.01	0.14	0.06	0.951	1.000	[-0.37, 0.39]
I only completed the task for money						
(Intercept)	3.36	0.10	33.82	< 0.001	< 0.001	[3.09, 3.63]
Monetary Incentive	0.06	0.11	0.61	0.541	1.000	[-0.22, 0.35]
UK	1.30	0.10	12.56	< 0.001	< 0.001	[1.02, 1.58]
Female	-0.01	0.07	-0.12	0.907	1.000	[-0.20, 0.18]
Non-Binary	0.24	0.48	0.51	0.610	1.000	[-1.04, 1.53]
Education	0.11	0.08	1.34	0.181	1.000	[-0.11, 0.34]
Age	-0.20	0.05	-4.23	< 0.001	< 0.001	[-0.32, -0.07]
Monetary Incentive × UK	-0.04	0.15	-0.28	0.778	1.000	[-0.45, 0.36]
I could have assessed more pictures if I'd tried harder						
(Intercept)	5.92	0.09	65.71	< 0.001	< 0.001	[5.68, 6.17]
Monetary Incentive	0.09	0.10	0.93	0.352	1.000	[-0.17, 0.35]
UK	-1.09	0.09	-11.60	< 0.001	< 0.001	[-1.35, -0.84]
Female	-0.13	0.06	-1.95	0.052	0.362	[-0.30, 0.05]
Non-Binary	0.41	0.43	0.94	0.348	1.000	[-0.76, 1.57]
Education	-0.13	0.08	-1.76	0.079	0.555	[-0.34, 0.07]
Age	0.05	0.04	1.17	0.241	1.000	[-0.06, 0.16]
Monetary Incentive × UK	-0.63	0.14	-4.62	< 0.001	< 0.001	[-0.99, -0.26]
I am satisfied with the amount of pay I received						
(Intercept)	6.03	0.06	93.73	< 0.001	< 0.001	[5.86, 6.21]
Monetary Incentive	0.01	0.07	0.07	0.941	1.000	[-0.18, 0.19]
UK	-0.35	0.07	-5.19	< 0.001	< 0.001	[-0.53, -0.17]

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
Female	0.01	0.05	0.31	0.753	1.000	[-0.11, 0.14]
Non-Binary	-0.41	0.31	-1.34	0.181	1.000	[-1.25, 0.42]
Education	-0.18	0.05	-3.24	0.001	0.008	[-0.32, -0.03]
Age	0.05	0.03	1.51	0.131	0.919	[-0.04, 0.13]
Monetary Incentive × UK	-0.09	0.10	-0.91	0.364	1.000	[-0.35, 0.17]

Results of linear regression models on the seven exploratory variables that participants completed after the image-rating task in Study 2a. Monetary Incentive is a binary variable for whether participants were in one of the monetary conditions. UK is a binary variable for whether participants were part of the UK sample. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Age is a continuous variable and is scaled. P_{Bonf} represents *p*-values after Bonferroni corrections for multiple comparisons.

Appendix Table S33a: **Exploratory Variables in Study 2b (Mexico and US Sample with Same Nominal Pay).**

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
I enjoyed completing the task						
(Intercept)	5.56	0.08	74.10	< 0.001	< 0.001	[5.36, 5.76]
Monetary Incentive	0.25	0.08	2.96	0.003	0.022	[0.02, 0.47]
US	-0.47	0.07	-6.28	< 0.001	< 0.001	[-0.67, -0.27]
Female	0.26	0.06	4.61	< 0.001	< 0.001	[0.11, 0.41]
Non-Binary	-0.34	0.18	-1.87	0.062	0.434	[-0.84, 0.15]
Education	-0.25	0.06	-4.11	< 0.001	< 0.001	[-0.41, -0.09]
Age	0.15	0.03	4.73	< 0.001	< 0.001	[0.06, 0.24]
Monetary Incentive × US	-0.16	0.12	-1.40	0.163	1.000	[-0.47, 0.15]
I am satisfied with how well I did on the task						
(Intercept)	6.31	0.05	116.09	< 0.001	< 0.001	[6.16, 6.45]
Monetary Incentive	0.08	0.06	1.32	0.186	1.000	[-0.08, 0.24]
US	-0.31	0.05	-5.70	< .001	< 0.001	[-0.45, -0.16]
Female	0.03	0.04	0.86	0.392	1.000	[-0.07, 0.14]
Non-Binary	-0.05	0.13	-0.36	0.720	1.000	[-0.41, 0.31]
Education	-0.01	0.04	-0.33	0.738	1.000	[-0.13, 0.10]
Age	0.06	0.02	2.64	0.008	0.059	[-0.00, 0.12]
Monetary Incentive × US	-0.11	0.08	-1.27	0.206	1.000	[-0.33, 0.12]
I believe that I helped others by completing the task						
(Intercept)	5.67	0.08	70.56	< 0.001	< 0.001	[5.46, 5.89]
Monetary Incentive	0.04	0.09	0.46	0.645	1.000	[-0.20, 0.28]
US	-0.74	0.08	-9.28	< 0.001	< 0.001	[-0.96, -0.53]
Female	0.07	0.06	1.18	0.238	1.000	[-0.09, 0.23]
Non-Binary	-0.34	0.20	-1.75	0.080	0.560	[-0.87, 0.19]
Education	-0.06	0.07	-0.93	0.353	1.000	[-0.24, 0.12]
Age	-0.02	0.03	-0.48	0.629	1.000	[-0.11, 0.08]
Monetary Incentive × US	-0.12	0.12	-0.95	0.342	1.000	[-0.45, 0.22]
Completing the task was boring						

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
(Intercept)	2.35	0.09	26.79	< 0.001	< 0.001	[2.11, 2.59]
Monetary Incentive	-0.12	0.10	-1.24	0.215	1.000	[-0.38, 0.14]
US	1.23	0.09	14.11	< 0.001	< 0.001	[0.99, 1.46]
Female	-0.29	0.07	-4.41	< 0.001	< 0.001	[-0.46, -0.11]
Non-Binary	0.27	0.21	1.27	0.204	1.000	[-0.31, 0.85]
Education	0.32	0.07	4.47	< 0.001	< 0.001	[0.13, 0.51]
Age	-0.28	0.04	-7.40	< 0.001	< 0.001	[-0.38, -0.18]
Monetary Incentive × US	0.08	0.14	0.61	0.540	1.000	[-0.28, 0.45]
I only completed the task for money						
(Intercept)	3.93	0.09	42.92	< 0.001	< 0.001	[3.68, 4.17]
Monetary Incentive	0.37	0.10	3.67	< 0.001	0.002	[0.10, 0.64]
US	1.25	0.09	13.79	< 0.001	< 0.001	[1.01, 1.50]
Female	-0.34	0.07	-4.95	< 0.001	< 0.001	[-0.52, -0.15]
Non-Binary	-0.37	0.22	-1.63	0.103	0.724	[-0.97, 0.24]
Education	0.14	0.07	1.95	0.051	0.360	[-0.06, 0.34]
Age	-0.19	0.04	-4.92	< 0.001	< 0.001	[-0.30, -0.09]
Monetary Incentive × US	-0.21	0.14	-1.50	0.135	0.942	[-0.59, 0.17]
I could have assessed more pictures if I'd tried harder						
(Intercept)	4.87	0.10	49.34	< 0.001	< 0.001	[4.61, 5.14]
Monetary Incentive	-0.07	0.11	-0.60	0.548	1.000	[-0.36, 0.23]
US	0.08	0.10	0.85	0.397	1.000	[-0.18, 0.35]
Female	-0.18	0.07	-2.39	0.017	0.118	[-0.37, 0.02]
Non-Binary	-0.79	0.24	-3.27	0.001	0.008	[-1.44, -0.14]
Education	0.10	0.08	1.24	0.217	1.000	[-0.12, 0.31]
Age	0.03	0.04	0.65	0.516	1.000	[-0.09, 0.14]
Monetary Incentive × US	-0.20	0.15	-1.30	0.194	1.000	[-0.61, 0.21]
I am satisfied with the amount of pay I received						
(Intercept)	5.69	0.06	88.92	< 0.001	< 0.001	[5.52, 5.86]
Monetary Incentive	0.13	0.07	1.79	0.074	0.518	[-0.06, 0.32]

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
US	0.23	0.06	3.55	< 0.001	0.003	[0.05, 0.40]
Female	0.02	0.05	0.43	0.670	1.000	[-0.11, 0.15]
Non-Binary	-0.42	0.16	-2.67	0.008	0.053	[-0.84, 0.00]
Education	-0.01	0.05	-0.28	0.779	1.000	[-0.15, 0.13]
Age	0.02	0.03	0.78	0.437	1.000	[-0.05, 0.09]
Monetary Incentive × US	-0.44	0.10	-4.50	< 0.001	< 0.001	[-0.71, -0.18]

Results of linear regression models on the seven exploratory variables that participants completed after the image-rating task in Study 2b for the comparison between samples in Mexico and the US (same nominal pay as in Mexico). Monetary Incentive is a binary variable for participants assigned to the monetary condition. US is a binary variable for participants in the sample from the US. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Age is a continuous variable and is scaled. P_{Bonf} represents *p*-values after Bonferroni corrections for multiple comparisons.

Appendix Table S33b: **Exploratory Variables in Study 2b (Mexico and US Sample with the Same Subjective Pay).**

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
I enjoyed completing the task						
(Intercept)	5.47	0.07	75.00	< 0.001	< 0.001	[5.28, 5.67]
Monetary Incentive	0.25	0.08	3.01	0.003	0.018	[0.03, 0.47]
US	-0.40	0.07	-5.58	< 0.001	< 0.001	[-0.60, -0.21]
Female	0.30	0.06	5.45	< 0.001	< 0.001	[0.15, 0.45]
Non-Binary	-0.12	0.18	-0.68	0.499	1.000	[-0.61, 0.36]
Education	-0.18	0.06	-3.08	0.002	0.015	[-0.34, -0.02]
Age	0.13	0.03	4.24	< 0.001	< 0.001	[0.05, 0.21]
Monetary Incentive × US	0.01	0.11	0.07	0.947	1.000	[-0.30, 0.31]
I am satisfied with how well I did on the task						
(Intercept)	6.24	0.05	116.35	< 0.001	< 0.001	[6.10, 6.39]
Monetary Incentive	0.08	0.06	1.34	0.182	1.000	[-0.08, 0.24]
US	-0.27	0.05	-5.04	< 0.001	< 0.001	[-0.41, -0.12]
Female	0.05	0.04	1.15	0.250	1.000	[-0.06, 0.16]
Non-Binary	0.03	0.13	0.25	0.804	1.000	[-0.32, 0.39]
Education	0.02	0.04	0.49	0.626	1.000	[-0.10, 0.14]
Age	0.02	0.02	0.86	0.389	1.000	[-0.04, 0.08]
Monetary Incentive × US	-0.01	0.08	-0.15	0.878	1.000	[-0.24, 0.21]
I believe that I helped others by completing the task						
(Intercept)	5.66	0.08	74.87	< 0.001	< 0.001	[5.46, 5.86]
Monetary Incentive	0.04	0.08	0.49	0.621	1.000	[-0.19, 0.27]
US	-0.67	0.07	-8.88	< 0.001	< 0.001	[-0.87, -0.46]
Female	0.09	0.06	1.52	0.128	0.895	[-0.07, 0.24]
Non-Binary	-0.26	0.19	-1.41	0.158	1.000	[-0.76, 0.24]
Education	-0.08	0.06	-1.33	0.183	1.000	[-0.25, 0.08]
Age	-0.04	0.03	-1.33	0.184	1.000	[-0.13, 0.04]
Monetary Incentive × US	0.03	0.12	0.30	0.766	1.000	[-0.28, 0.35]
Completing the task was boring						

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
(Intercept)	2.39	0.09	27.94	< 0.001	< 0.001	[2.16, 2.62]
Monetary Incentive	-0.12	0.10	-1.23	0.219	1.000	[-0.38, 0.14]
US	1.05	0.08	12.35	< 0.001	< 0.001	[0.82, 1.28]
Female	-0.27	0.06	-4.14	< 0.001	< 0.001	[-0.44, -0.09]
Non-Binary	0.02	0.21	0.09	0.931	1.000	[-0.55, 0.59]
Education	0.32	0.07	4.57	< 0.001	< 0.001	[0.13, 0.51]
Age	-0.20	0.04	-5.44	< 0.001	< 0.001	[-0.30, -0.10]
Monetary Incentive × US	-0.06	0.13	-0.45	0.653	1.000	[-0.42, 0.30]
I only completed the task for money						
(Intercept)	3.97	0.09	42.84	< 0.001	< 0.001	[3.72, 4.22]
Monetary Incentive	0.37	0.10	3.60	< 0.001	0.002	[0.09, 0.65]
US	1.12	0.09	12.20	< 0.001	< 0.001	[0.87, 1.37]
Female	-0.33	0.07	-4.72	< 0.001	< 0.001	[-0.52, -0.14]
Non-Binary	-0.62	0.23	-2.71	0.007	0.047	[-1.23, -0.00]
Education	0.14	0.08	1.85	0.065	0.455	[-0.06, 0.34]
Age	-0.13	0.04	-3.36	0.001	0.006	[-0.24, -0.03]
Monetary Incentive × US	-0.48	0.14	-3.33	0.001	0.006	[-0.87, -0.09]
I could have assessed more pictures if I'd tried harder						
(Intercept)	4.84	0.10	49.63	< 0.001	< 0.001	[4.58, 5.11]
Monetary Incentive	-0.06	0.11	-0.55	0.584	1.000	[-0.35, 0.23]
US	0.08	0.10	0.82	0.412	1.000	[-0.18, 0.34]
Female	-0.08	0.07	-1.03	0.305	1.000	[-0.27, 0.12]
Non-Binary	-0.78	0.24	-3.26	< 0.001	0.008	[-1.43, -0.14]
Education	0.11	0.08	1.32	0.186	1.000	[-0.11, 0.32]
Age	0.07	0.04	1.69	0.090	0.632	[-0.04, 0.18]
Monetary Incentive × US	-0.40	0.15	-2.65	0.008	0.056	[-0.81, 0.01]
I am satisfied with the amount of pay I received						
(Intercept)	5.70	0.06	91.98	< 0.001	< 0.001	[5.53, 5.86]
Monetary Incentive	0.13	0.07	1.86	0.064	0.446	[-0.06, 0.32]

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
US	0.27	0.06	4.37	< 0.001	< 0.001	[0.10, 0.43]
Female	0.10	0.05	2.07	0.039	0.272	[-0.03, 0.22]
Non-Binary	-0.16	0.15	-1.08	0.282	1.000	[-0.58, 0.25]
Education	-0.08	0.05	-1.60	0.111	0.774	[-0.22, 0.06]
Age	0.02	0.03	0.82	0.411	1.000	[-0.05, 0.09]
Monetary Incentive × US	-0.43	0.10	-4.47	< 0.001	< 0.001	[-0.69, -0.17]

Results of linear regression models on the seven exploratory variables that participants completed after the image-rating task in Study 2b for the comparison between samples in Mexico and the US (same subjective pay as in Mexico). Monetary Incentive is a binary variable for participants assigned to the monetary condition. US is a binary variable for participants in the sample from the US. Female and Non-Binary are categorical variables for gender. Education is a binary variable specifying people who had either completed or were pursuing studies at the undergraduate level or higher at the time of the experiment. Age is a continuous variable and is scaled. P_{Bonf} represents *p*-values after Bonferroni corrections for multiple comparisons.

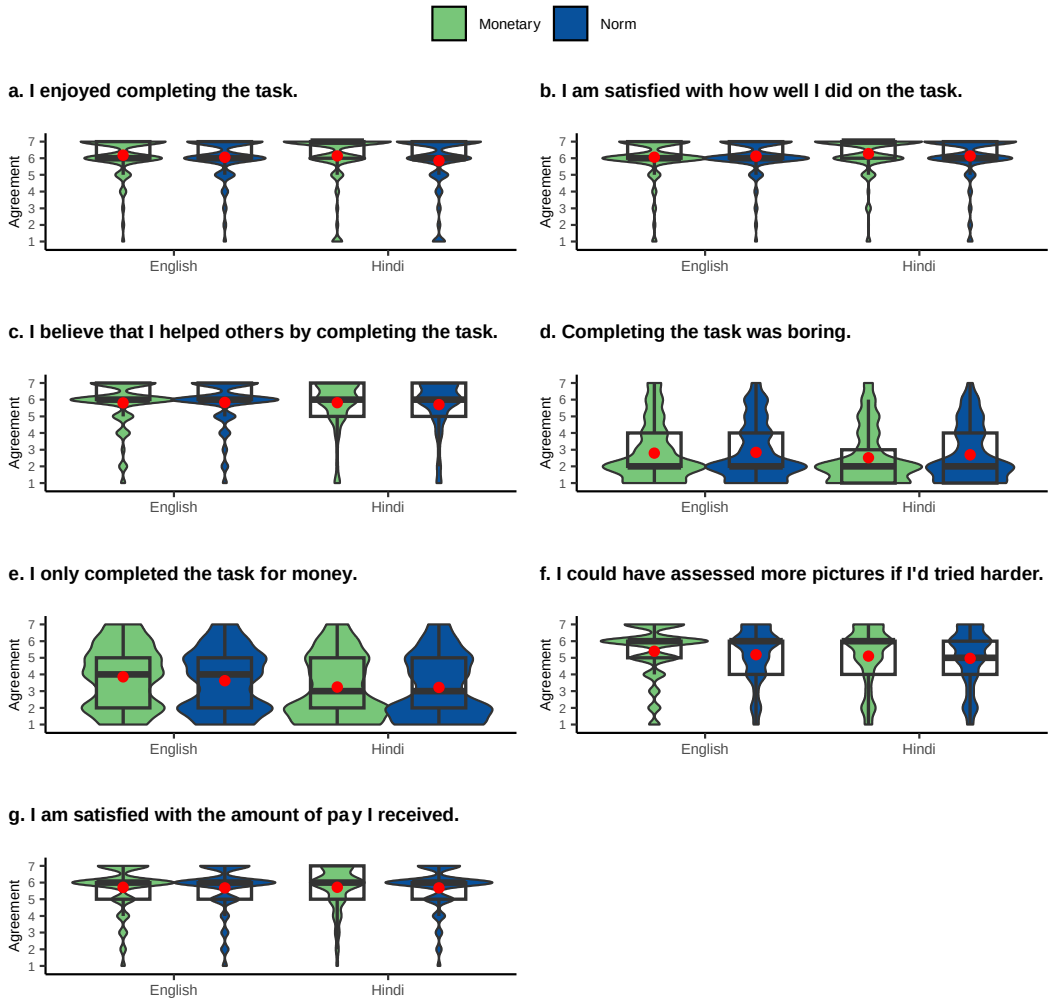
Appendix Table S34: **Exploratory Variables in Study 4.**

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
I enjoyed completing the task						
(Intercept)	5.85	0.06	98.02	< 0.001	< 0.001	[5.69, 6.01]
Monetary Incentive	0.29	0.08	3.41	0.001	0.005	[0.06, 0.51]
English	0.21	0.08	2.44	0.015	0.104	[-0.02, 0.43]
Monetary Incentive × English	-0.17	0.12	-1.38	0.167	1.000	[-0.49, 0.16]
I am satisfied with how well I did on the task						
(Intercept)	6.14	0.05	122.27	< 0.001	< 0.001	[6.00, 6.27]
Monetary Incentive	0.13	0.07	1.89	0.059	0.410	[-0.06, 0.33]
English	-0.01	0.07	-0.21	0.833	1.000	[-0.21, 0.18]
Monetary Incentive × English	-0.20	0.10	-1.98	0.048	0.337	[-0.47, 0.07]
I believe that I helped others by completing the task						
(Intercept)	5.70	0.06	91.22	< 0.001	< 0.001	[5.53, 5.87]
Monetary Incentive	0.12	0.09	1.34	0.180	1.000	[-0.12, 0.36]
English	0.14	0.09	1.57	0.117	0.822	[-0.10, 0.38]
Monetary Incentive × English	-0.15	0.13	-1.21	0.226	1.000	[-0.49, 0.19]
Completing the task was boring						
(Intercept)	2.69	0.07	36.23	< 0.001	< 0.001	[2.49, 2.89]
Monetary Incentive	-0.17	0.11	-1.64	0.100	0.703	[-0.46, 0.11]
English	0.15	0.11	1.41	0.160	1.000	[-0.14, 0.43]
Monetary Incentive × English	0.13	0.15	0.84	0.402	1.000	[-0.28, 0.53]
I only completed the task for money						
(Intercept)	3.22	0.08	39.62	< .001	< 0.001	[3.00, 3.44]
Monetary Incentive	0.02	0.11	0.21	0.831	1.000	[-0.29, 0.33]
English	0.40	0.12	3.49	0.001	0.004	[0.09, 0.71]
Monetary Incentive × English	0.21	0.16	1.30	0.195	1.000	[-0.23, 0.65]
I could have assessed more pictures if I'd tried harder						
(Intercept)	4.97	0.07	67.40	< 0.001	< 0.001	[4.77, 5.17]

	b	*SE*	*t*	*P*	P_{Bonf}	99.3% CI
Monetary Incentive	0.13	0.10	1.30	0.195	1.000	[-0.15, 0.42]
English	0.22	0.10	2.11	0.035	0.242	[-0.06, 0.50]
Monetary Incentive × English	0.07	0.15	0.50	0.617	1.000	[-0.33, 0.47]
I am satisfied with the amount of pay I received						
(Intercept)	5.68	0.06	99.36	< 0.001	< 0.001	[5.53, 5.83]
Monetary Incentive	0.03	0.08	0.43	0.670	1.000	[-0.18, 0.25]
English	0.00	0.08	0.01	0.980	1.000	[-0.22, 0.22]
Monetary Incentive × English	0.00	0.11	-0.03	0.980	1.000	[-0.31, 0.31]

Results of linear regression models on the seven exploratory variables that participants completed after the image-rating task in Study 4. Monetary Incentive is a binary variable for being assigned to the monetary condition. English is a binary variable for those who took the survey in English (as opposed to in Hindi). P_{Bonf} represents *p*-values after Bonferroni corrections for multiple comparisons.

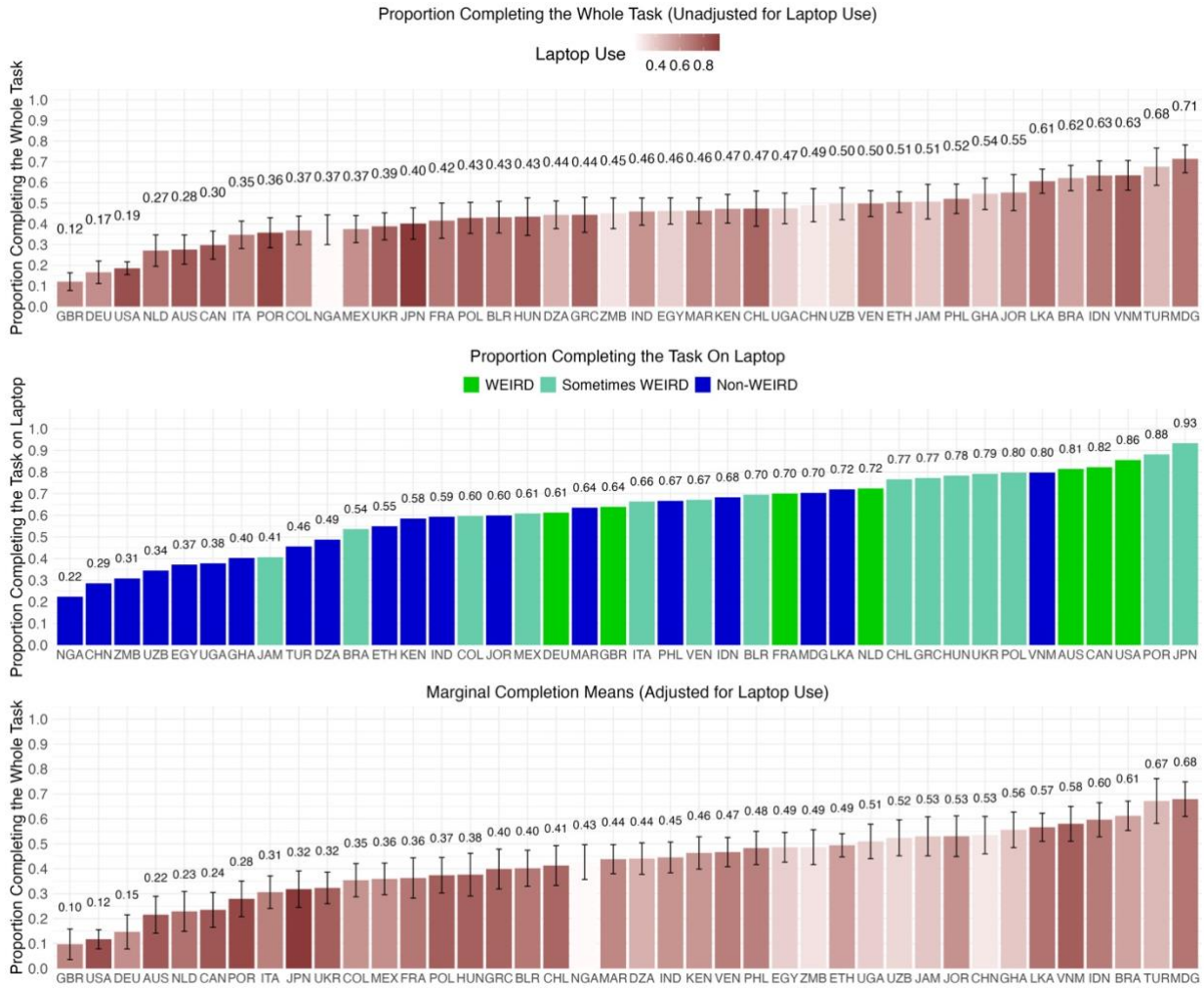
Fig. S4: Exploratory Variables in Study 4, by Language and Incentive Type.



Effects of the monetary condition (green) and the norm condition (blue), by assigned language (English or Hindi), on the seven exploratory variables that participants in India (N = 2,065 participants recruited on Facebook) completed after the image-rating task in Study 4. The graph displays the central tendency and distribution of effort by incentive condition. The black line within each box represents a median, and the red dot shows a mean. Upper and lower bounds show the third and the first quartile, respectively. The whiskers represent the 1.5 times the interquartile range, with black points showing observations outside of this range. The width of each violin corresponds to the frequency of observations at any given number of images rated on the y-axis. Appendix Table S*16* reports the results of the statistical tests.

RATES OF LAPTOP USE IN STUDY 5A AND COMPARISON OF RAW AND MARGINAL MEANS

***Fig. S5*: Raw and Marginal Proportions Completed Adjusted for Prevalence of Laptop Use.**



Comparison between raw (top) and marginal means (bottom) task completion rates based on the proportion of participants in each country who completed the study on a laptop or personal in Study 5a (N = 7,823 independent participants recruited on seven crowd-sourcing platforms). Laptop use was self-reported by participants. Marginal means adjust for device type and display the results as if the samples had the same composition of phone and laptop users by fitting a linear model to estimate the average outcomes for each group while holding the proportion working on a laptop constant, allowing for more accurate comparisons of means across countries. Regression results are reported in the results section and in the legend for Additional Fig. 6. Error bars are 95% CIs.

STUDY 5B

In Study 2, we zeroed in on the mechanism of relational understandings of contracts by adding a framing that encouraged people to think about the work setup as more relational and less legalistically spelled out. To make the contract more relational and less legalistic, we added a small statement that made the pay sound more ambiguous. We told workers that "their pay will be finalized after the task." This could encourage a relational understanding because it implies that we would look at how hard they worked to determine their pay.

Our main predictions concerned the main effect of culture would remain the same as in Study 5a. That is, we thought that Western participants would be less likely to complete the entire task than their "non-WEIRD" counterparts, just like in Study 5A. In other words, we expected that regardless of whether the instructions are explicit about the lack of remuneration or are ambiguous but do not guarantee it, "WEIRD"—compared to "non-WEIRD"—workers would be less likely to exert effort beyond the minimum required. We also expected to observe a main effect of condition, whereby people across cultures would show higher completion rates when we hinted that their effort might lead to more earnings.

With less certainty, we predicted that an increase in effort in the ambiguous condition (relative to the "explicitly no pay" condition) would be higher in the West. We thought that encouraging people to think of the contract as relational should make a larger difference in Western cultures because people in Western cultures by default are not thinking of contracts as relational. On the other hand, if relational contracting is the more prevalent schema for non-Western workers, they might be more motivated to do more than the contractual minimum even without the promise of additional rewards. Therefore, non-Western workers might not perceive the ambiguity inherent in this condition as fundamentally distinct from the condition where we

clearly stated that higher effort would not lead to higher pay. Second, we though that Western workers might be simply more motivated by the possibility of earning more money: if receiving additional pay is more motivating in the West, we should see that the possibility of earning money similarly increases effort more there as in our Studies 1–3. Ambiguous pay can be seen as a monetary incentive of unspecified size and with discounted expected value due to the uncertainty that one will actually receive it.

On other hand, we thought that it was possible that Western workers would be less likely to believe that our remark—"pay will be finalized after the task"—could actually lead to personal benefits. They might trust such ambiguous suggestions less: if it is not clearly written in the contract, it is not worth fighting for. They might also be more risk-averse and less predisposed to pay the cost of effort.

Results

We first analyzed the sample from Study 5B (N = 7,578) individually, that is separately from our sample in Study 5A. The mixed-effects logistic regression analysis showed a significant negative linear effect of the "WEIRD" country classification on the probability of completing the whole task, $b = -0.852$, $z = -5.852$, $P < 0.001$, 95% CI -1.137 to -0.567, indicating that participants from more "WEIRD" countries were less likely to complete the entire transcription (Fig. S5). The quadratic term for the "WEIRD" classification was not significant, $b = -0.023$, $z = -0.233$, $P = 0.815$ 95% CI -0.220 to 0.173, suggesting that the decline in completion rates was primarily driven by the linear effect.

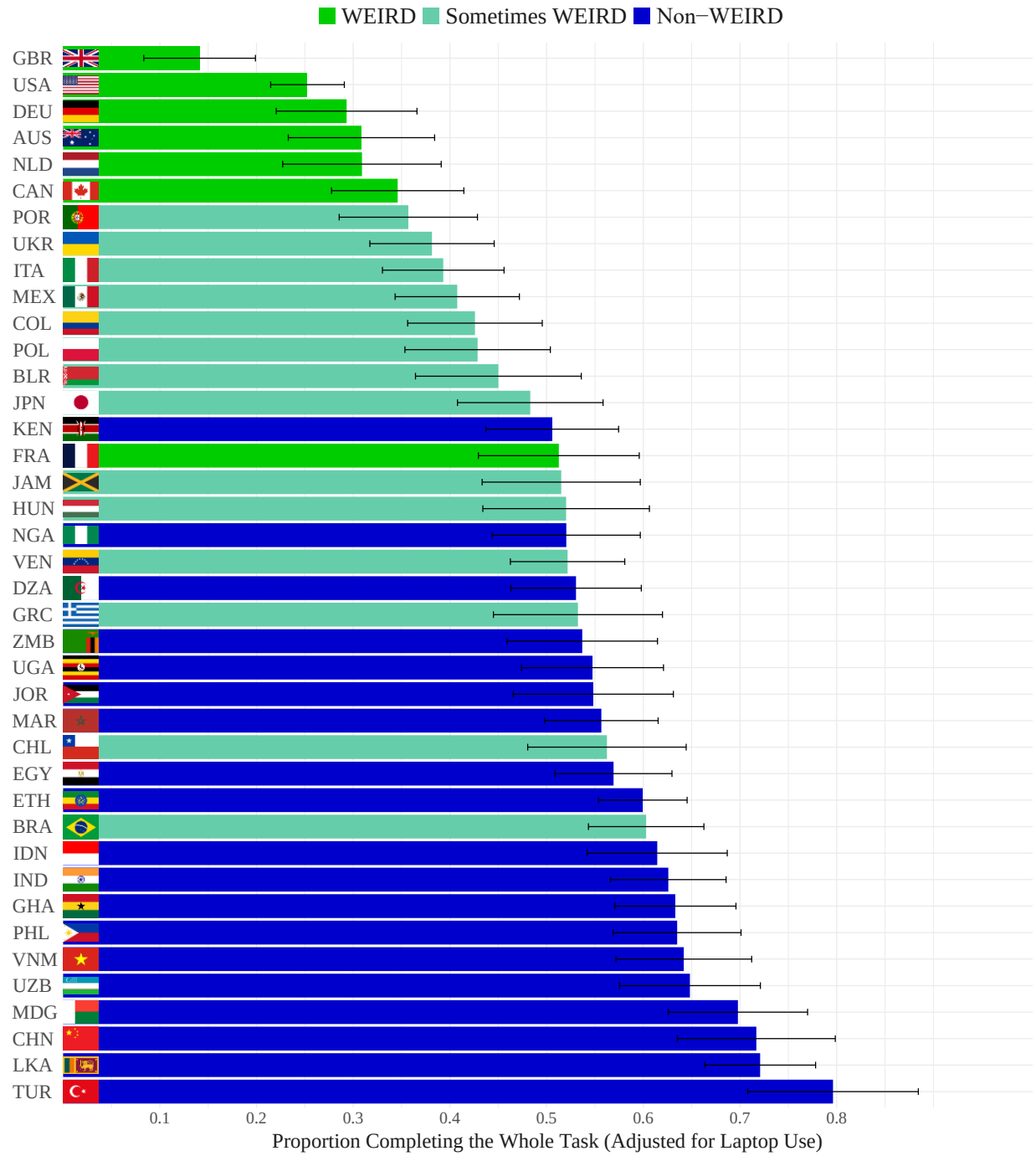
When we analyzed the combined sample across Studies 5A and 5B, the mixed-effects logistic regression analysis showed that participants across countries in the ambiguous condition, where the possibility of additional pay was implied but not explicit, were significantly more

likely to complete the whole task compared to those in the explicitly-no-pay condition, b = 0.436, t = 10.565, P < 0.001, 95% CI 0.355 to 0.517 (Fig. S6).

The linear term for "WEIRD" shows a significant negative effect, indicating that participants from "WEIRD" countries were less likely to complete the whole task, b = -0.996, t = -6.777, P < 0.001, 95% CI -1.284 to -0.708 (Fig. S7). The quadratic term for "WEIRD" also showed a significant negative effect, b = -0.239, t = -2.061, P = 0.039, 95% CI -0.466 to -0.012, suggesting that the decline in completion rates along the "WEIRD" classification was not linear. The interaction between condition (ambiguous) and the linear term for "WEIRD" was not significant, b = 0.133, t = 1.784, P = 0.074, 95% CI -0.013 to 0.280. The interaction between condition (ambiguous) and the quadratic term for "WEIRD" was significant, suggesting that the effect of the ambiguous condition on task completion was more pronounced in "WEIRD" or "Partly (Sometimes) WEIRD" cultures, b = 0.168, t = 2.467, P = 0.014, 95% CI 0.035 to 0.302.

Therefore, in Study 5B, we found that participants in "WEIRD" countries were less likely to complete the task even when extra pay was implied but not explicitly promised. In addition, the difference in completion rates between "WEIRD" and "non-WEIRD" countries was smaller in this ambiguous condition compared to the "Explicitly No Pay" condition in Study 5A, suggesting that contractual ambiguity of pay might have a larger effect on effort in Western countries.

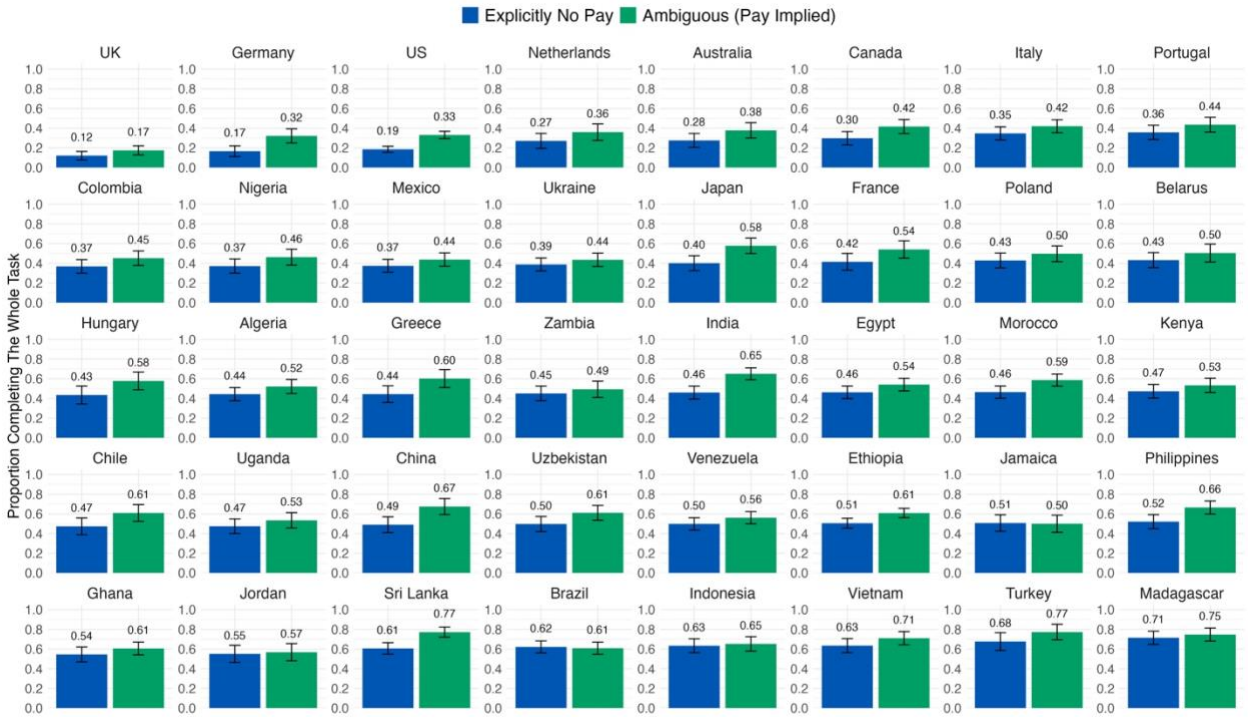
Fig. S6: **Study 5b, Proportion of Participants per Country Who Complete the Whole Task Under Ambiguous Pay.**



Proportion of participants in Study 5b (N = 7,578 independent participants recruited on seven crowd-sourcing platforms) who complete the entire data entry task when the monetary reward for doing more than the contractual minimum is ambiguous. The figure presents the marginal means for completion rates adjusted for laptop use,

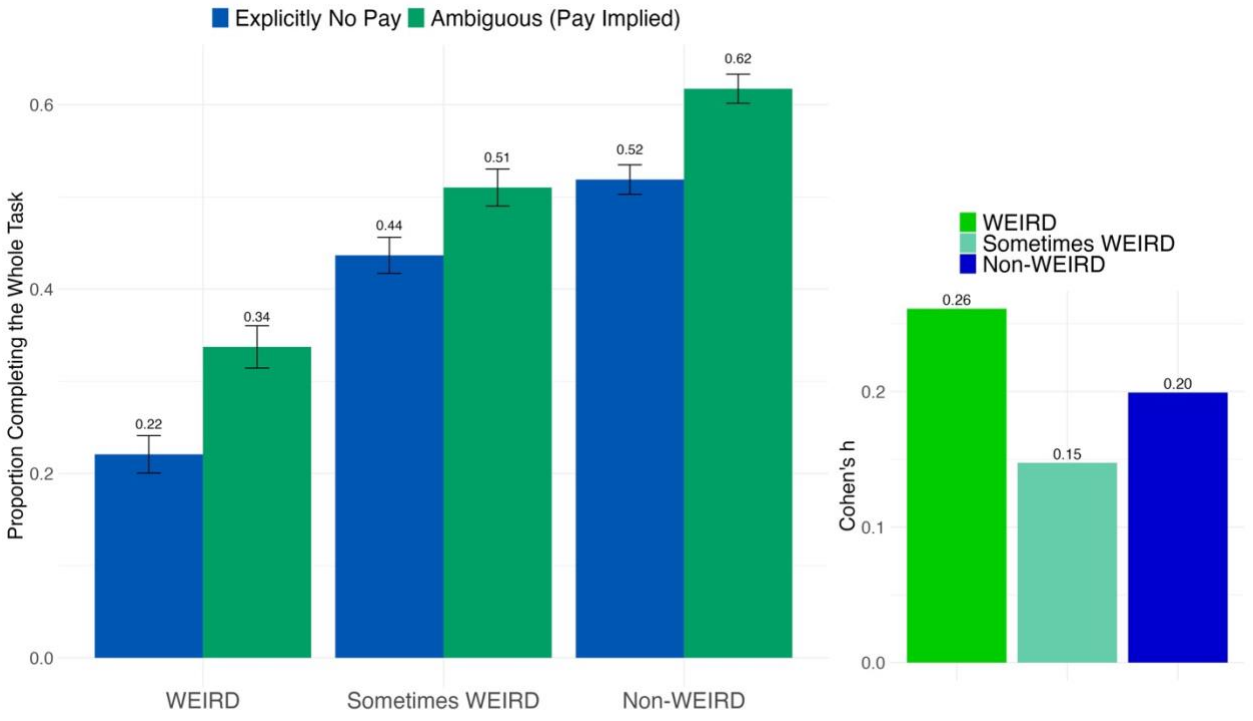
colored by “WEIRD” country classification adapted from Talhelm et al (under review). The mixed-effects logistic regression analysis showed a significant negative linear effect of the “WEIRD” country classification on the probability of completing the whole task, $b = -0.852$, $z = -5.852$, $P < 0.001$, 95% CI -1.137 to -0.567, indicating that participants from more “WEIRD” countries are less likely to complete the entire transcription. The quadratic term for the “WEIRD” classification was not significant, $b = -0.023$, $z = -0.233$, $P = 0.815$, 95% CI -0.220 to 0.173, suggesting that the decline in completion rates was primarily driven by the linear effect. Error bars are 95% CIs.

Fig. S7a: **Proportion Completing the Whole Task in Studies 5A (Explicitly No Pay) and 5B (Ambiguous, Pay Implied), by Country.**



Proportions of participants in each country (N = 15,485 independent participants recruited on crowd-sourcing platforms) completing the whole task in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms, in blue), where they were explicitly told that they will not earn more money for working beyond the 5-minute minimum, and in Study 5b (N = 7,578 independent participants recruited on crowd-sourcing platforms, in green), where the lack of additional pay was not explicit and instead participants were told that their pay would be "finalized after the task." The countries are ordered from lowest to highest completion rate in Study 5a (Explicitly No Pay). Error bars represent 95% CIs.

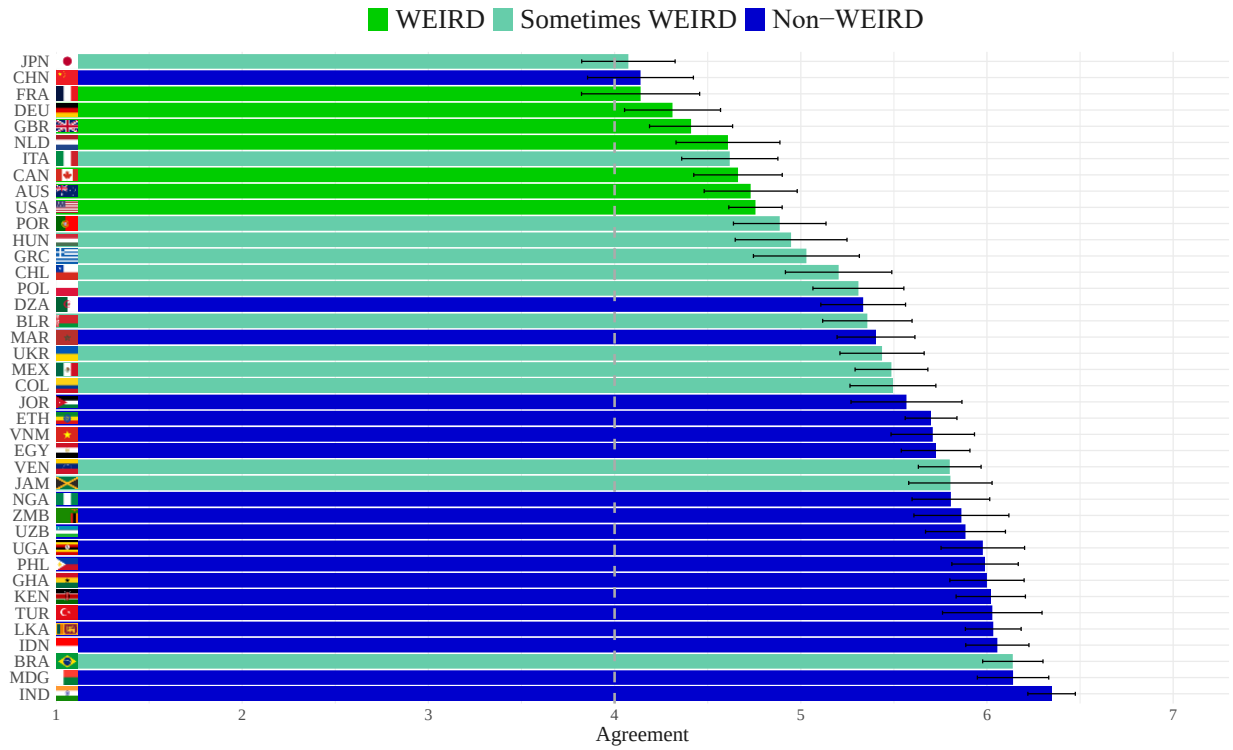
Fig. S7b: **Proportion Completing the Whole Task in Studies 5A (Explicitly No Pay) and 5B (Ambiguous, Pay Implied), by "WEIRD" Classification.**



Proportions of participants (N = 15,485 independent participants recruited on crowd-sourcing platforms) completing the whole task, per "WEIRD" classification, in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms, in blue), where they were explicitly told that they will not earn more money for working beyond the 5-minute minimum, and in Study 5b (N = 7,578 independent participants recruited on crowd-sourcing platforms, in green), where the lack of additional pay was not explicit and instead participants were told that their pay would be "finalized after the task." The mixed-effects logistic regression analysis showed that participants across countries in the ambiguous condition, where the possibility of additional pay was implied but not explicit, are significantly more likely to complete the whole task compared to those in the explicitly no pay condition, b = 0.436, t = 10.565, P < 0.001, 95% CI 0.355 to 0.517. The linear term for "WEIRD" shows a significant negative effect, indicating that participants from more "WEIRD" countries are less likely to complete the whole task, b = -0.996, t = -6.777, P < 0.001, 95% CI -1.284 to -0.708. The quadratic term for "WEIRD" also shows a significant negative effect, b = -0.239, t = -2.061, P = 0.039, 95% CI -0.466 to -0.012, suggesting that the effect of the "WEIRD" classification was not the same along the "WEIRD" continuum. The interaction between condition (ambiguous) and the linear term for "WEIRD" is not significant, b = 0.133, t = 1.784, P = 0.074, 95% CI -0.013 to 0.280. The interaction between condition (ambiguous) and the quadratic term for "WEIRD" is significant, suggesting that the effect of the ambiguous condition on task completion was more pronounced in "WEIRD" or "Partly (Sometimes) WEIRD" cultures, b = 0.168, t = 2.467, P = 0.014, 95% CI 0.035 to 0.302.

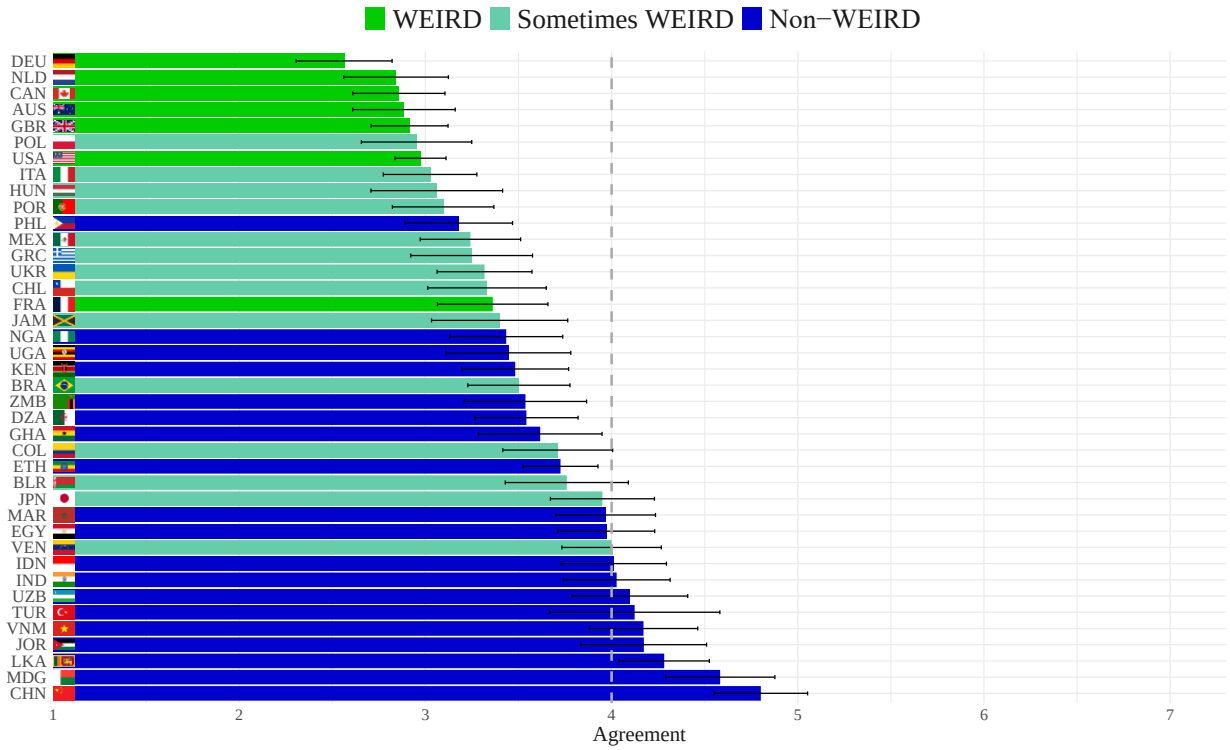
EXPLORATORY VARIABLES IN STUDIES 5A AND 5B

Fig. S8: Self-Reported Task Enjoyment in Study 5A, by Country.



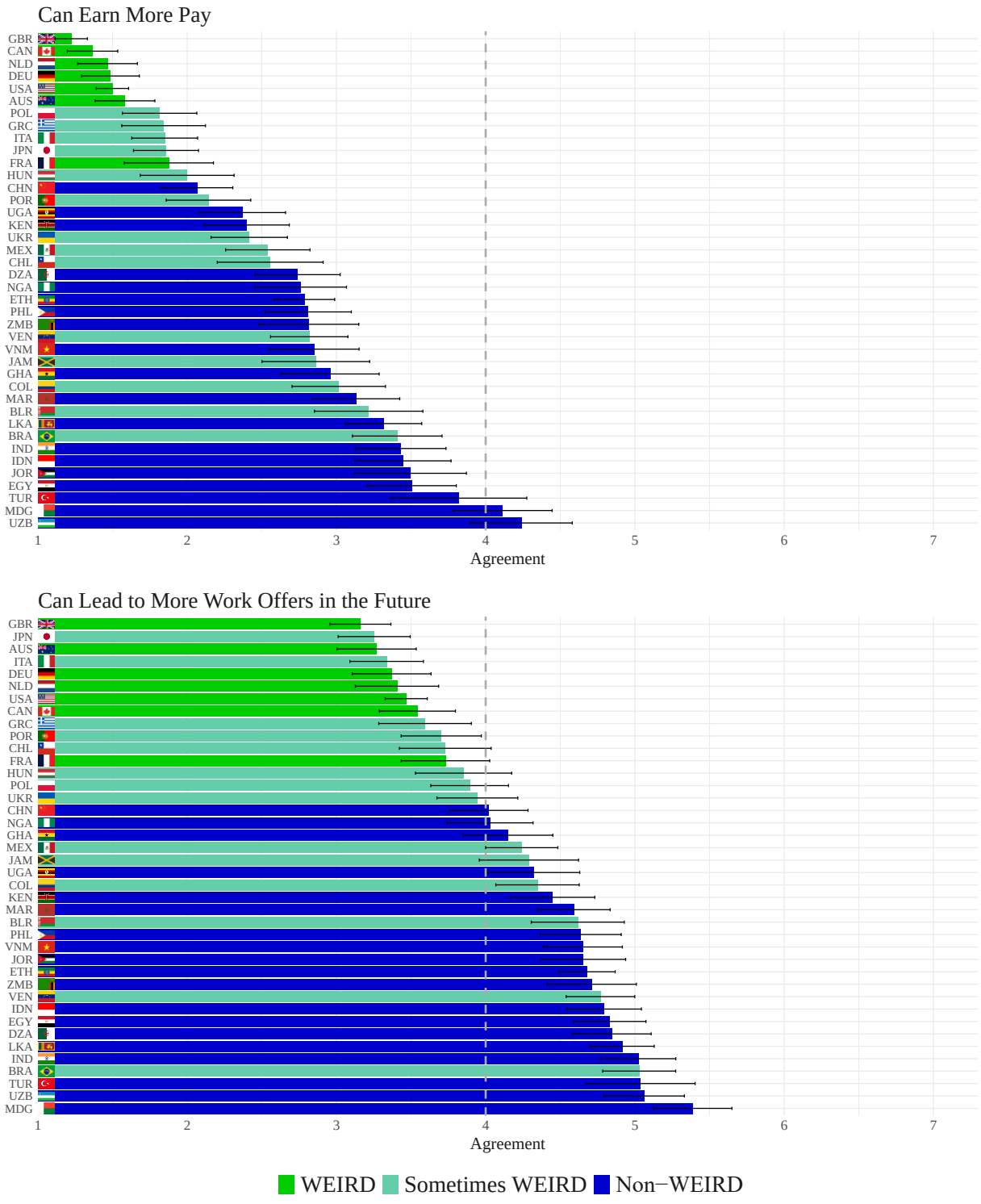
The regression output for the mixed-effects regression model on agreement with the item "I enjoyed completing the task" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms) in 40 countries. The bars are colored according to whether they country was categorized as "WEIRD" (bright green), "Partly (Sometimes) WEIRD" (in aquamarine), and "non-WEIRD" (in blue). The countries within each panel are ordered from lowest to highest agreement. Participants who completed the whole task are significantly more likely to report higher task enjoyment, b = 0.476, t = 16.819, P < 0.001, 95% CI 0.420 to 0.531. The linear term for "WEIRD" shows a significant negative effect, indicating that participants from "WEIRD" countries are less likely to enjoy the task, b = -0.354, t = -4.797, P < 0.001, 95% CI -0.498 to -0.209. Conversely, the quadratic term for "WEIRD" was not significant, b = -0.064, t = -1.304, P = 0.199, 95% CI -0.159 to 0.032, suggesting that the quadratic effect does not significantly influence task enjoyment. The interaction between task completion and the linear term for "WEIRD" was not significant, indicating that among people who complete the whole task, task enjoyment was not significantly different for those from "WEIRD" countries, b = 0.078, t = 1.547, P = 0.122, 95% CI -0.021 to 0.176. The interaction between task completion and the quadratic term for "WEIRD" was significant, suggesting that among people who complete the task enjoyment increases non-gradually across people in "WEIRD" and "Partly (Sometimes) WEIRD" cultures, b = 0.118, t = 2.559, P = 0.011, 95% CI 0.028 to 0.208. Error bars represent 95% CIs.

Fig. S9: **Perceptions of Researchers' Expectations in Study 5A, by Country.**



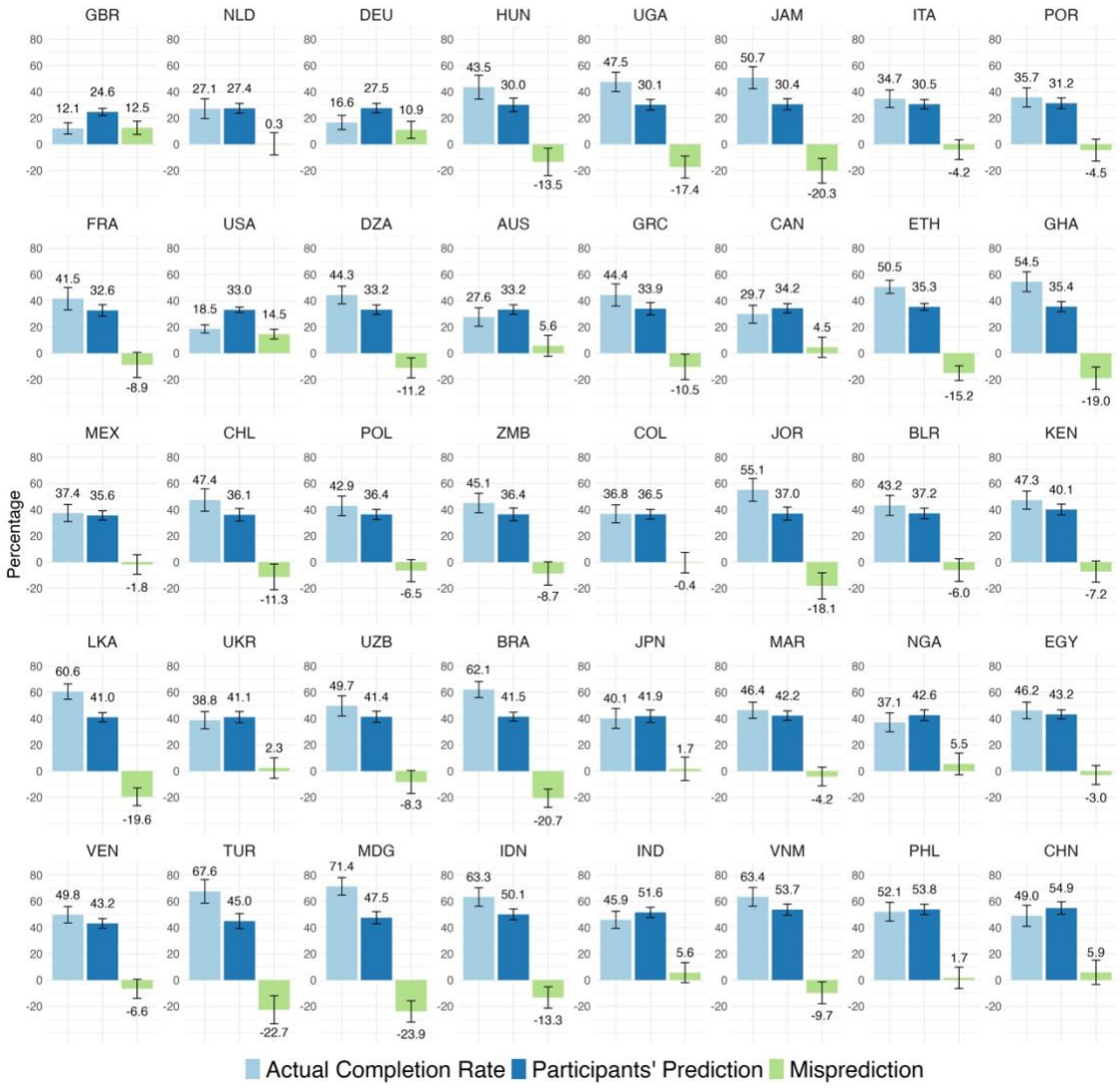
Country-level means for the item "The researchers wanted me to work for longer than 5 minutes" in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms) in 40 countries. The bars are colored according to whether they country was categorized as "WEIRD" (bright green), "Partly (Sometimes) WEIRD" (in aquamarine), and "non-WEIRD" (in blue). The countries within each panel are ordered from lowest to highest agreement. Participants who completed the whole task are significantly more likely to perceive that the researchers wanted them to work for more than 5 minutes, b = 0.349, t = 11.807, P < 0.001, 95% CI 0.291 to 0.407 (Appendix Table S25). The linear term for "WEIRD" was not significant, indicating that participants from "WEIRD" countries do not significantly differ in their perceptions of researchers' expectations, b = -0.014, t = -0.223, P = 0.824, 95% CI -0.135 to 0.107. The quadratic term for "WEIRD" was also not significant, b = 0.068, t = 1.646, P = 0.109, 95% CI -0.013 to 0.148. The interaction between task completion and the linear term for "WEIRD" was significant, indicating that among people who complete the whole task, perceptions that researchers wanted them to work more than 5 minutes were less pronounced for those from "WEIRD" countries, b = -0.150, t = -2.779, P = 0.008, 95% CI -0.256 to -0.044. Similarly, the interaction between task completion and the quadratic term for "WEIRD" was significant, suggesting a non-linear effect along the "WEIRD" classification, b = -0.137, t = -2.840, P = 0.007, 95% CI -0.231 to -0.042. Error bars represent 95% CIs.

***Fig. S10*: Expectations of Higher Pay and Future Work Opportunities for Task Completion in Study 5A, by Country.**



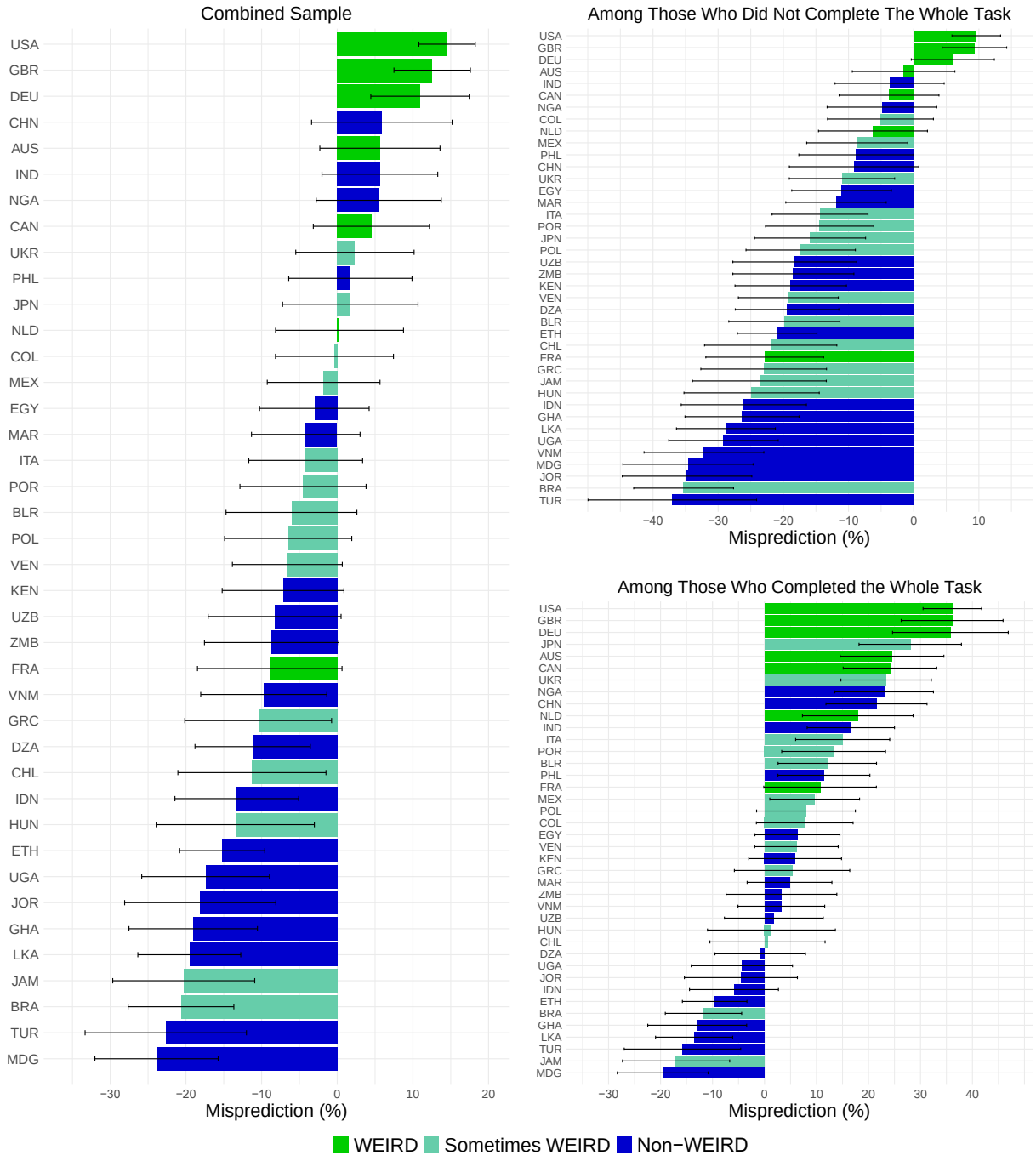
Country-level means for the items “Working for longer than 5 minutes could have earned me more pay” (top) and “Working for longer than 5 minutes could encourage the researchers to give me more work in the future” (bottom) in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms) in 40 countries. The bars are colored according to whether they country was categorized as “WEIRD” (bright green), “Partly (Sometimes) WEIRD” (in aquamarine), and “non-WEIRD” (in blue). The countries within each panel are ordered from lowest to highest agreement. Regression output is presented in Appendix Tables S22 and S23.

Fig. S11: **Actual Completion Rates, and Participants' Predictions and Mispredictions of Effort in Study 5A, by Country.**



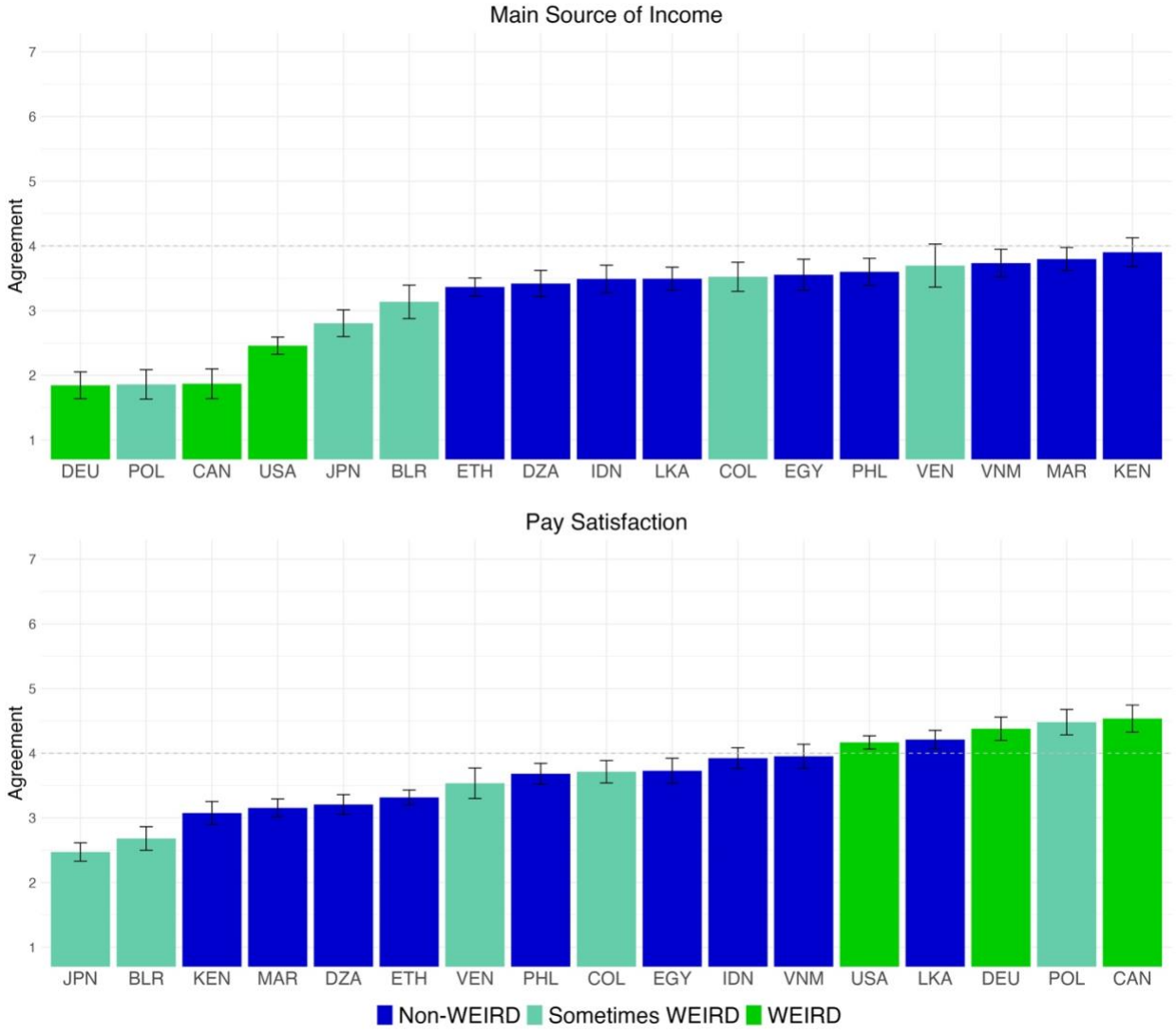
Actual mean rates per country of completing the whole task (light blue), participants' mean predicted completion rates in their country (dark blue), and mean mispredictions, or differences between predicted and actual completion rates, (light green) in each of the 40 countries in Study 5a (N = 7,823 independent participants recruited on crowd-sourcing platforms). The countries are ordered from lowest to highest predicted completion rate. Error bars represent 95% CIs.

Fig. S12: **Mispredictions of Effort in Study 5A, by Country and Task Completion.**



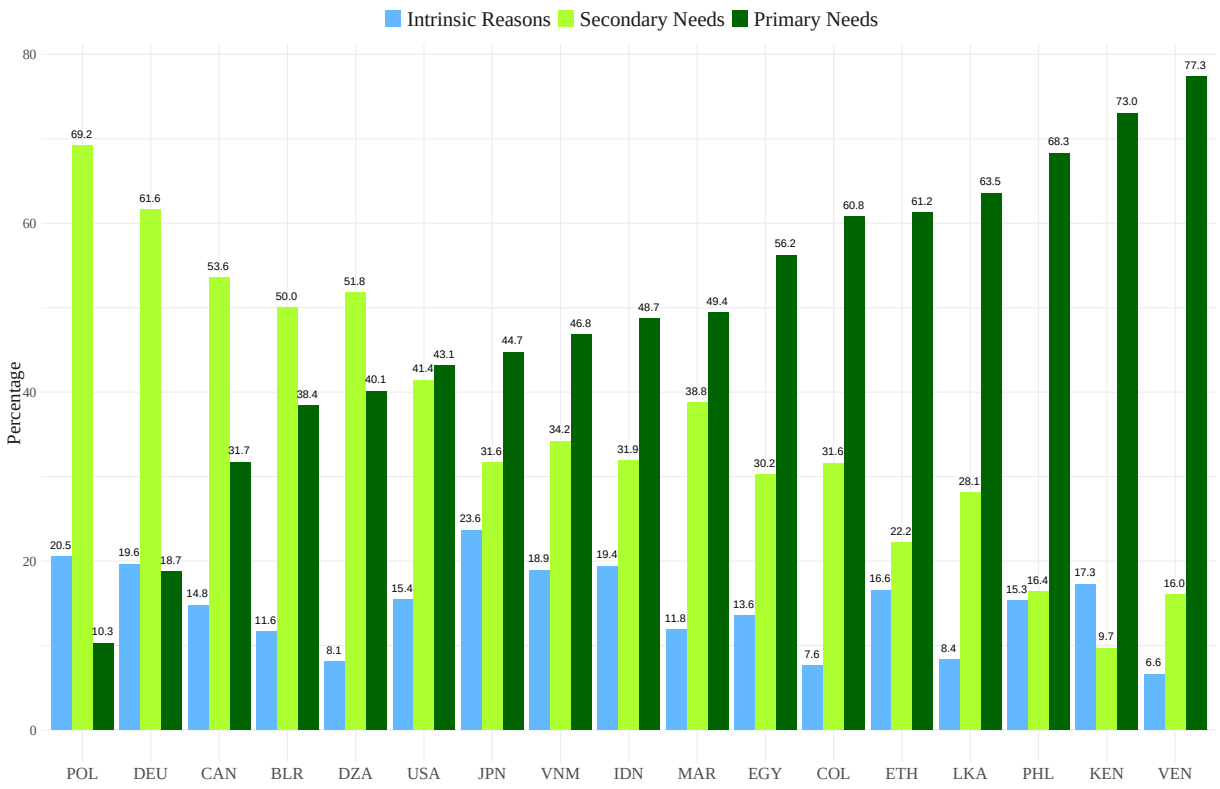
Mean mispredictions, or the difference between predicted and actual completion rates in each of the 40 countries (*N* = 7,823 independent participants recruited on crowd-sourcing platforms) on the data entry task in Study 5a. The bars are colored according to whether they country was categorized as "WEIRD" (bright green), "Partly (Sometimes) WEIRD" (in aquamarine), and "non-WEIRD" (in blue). The graph on the left shows the combined sample. The two graphs on the right only include participants who did not (top) versus did (bottom) complete the entire data task (top). The countries within each panel are ordered from largest positive to largest negative misprediction. Error bars represent 95% CIs.

Fig. S13: **Centrality of Crowd-Sourcing Income and Pay Satisfaction with Wages.**



Mean agreement with items "Crowd-sourcing is my main source of income" (top) and "I am satisfied with how much money I earn on (Platform)." (Platform) was replaced by a specific crowd-sourcing platform in each sample. The items were measured on 1–7 Likert scales. 8,439 independent participants across Studies 5a and 5b completed the additional set of survey measures that included these items. The countries are ordered from lowest to highest agreement with each of the items. Error bars represent 95% confidence intervals.

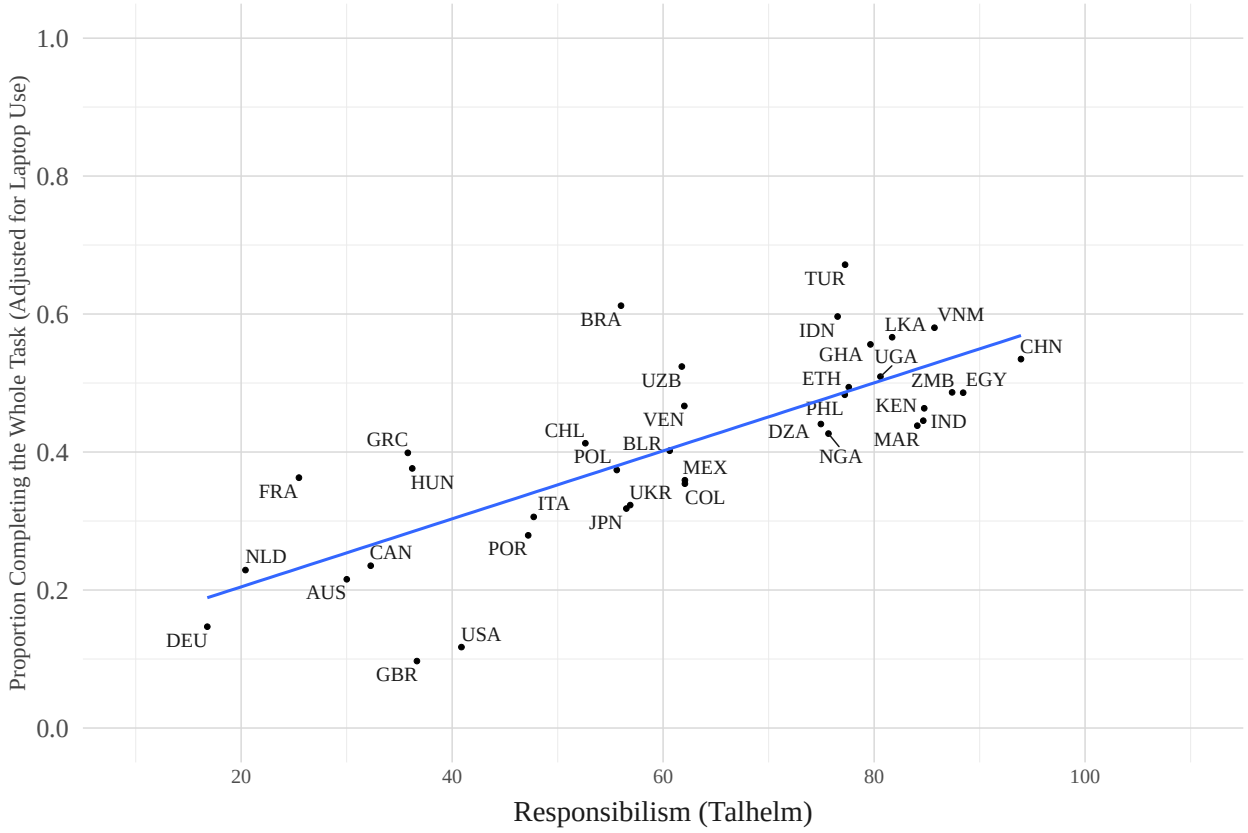
Fig. S14: **Main Reported Motivation for Doing Crowd-Source Work Among Participants in Studies 5A and 5B.**



Percentages of participants in Studies 5a and 5b (combined N = 15,485 independent participants recruited on crowd-sourcing platforms) in each country who selected one of the three main reason categories for doing crowd-source work. The possible options were 'To earn money for food, rent, or bills,' which we renamed 'Primary Needs;' 'To earn money for hobbies, gadgets, or going out,' which we renamed 'Secondary Needs;' and five answer choices that we merged into a category 'Intrinsic Reasons:' 'To kill time,' 'Because the tasks are fun,' 'To benefit society,' 'To learn new things,' and 'To share my opinions with someone who cares about what I think. This figure includes countries where more than 180 participants were asked this question, as the section that included this question was added after the initial round of data collection. The countries are ordered from lowest to highest motivation to earn money for primary needs.

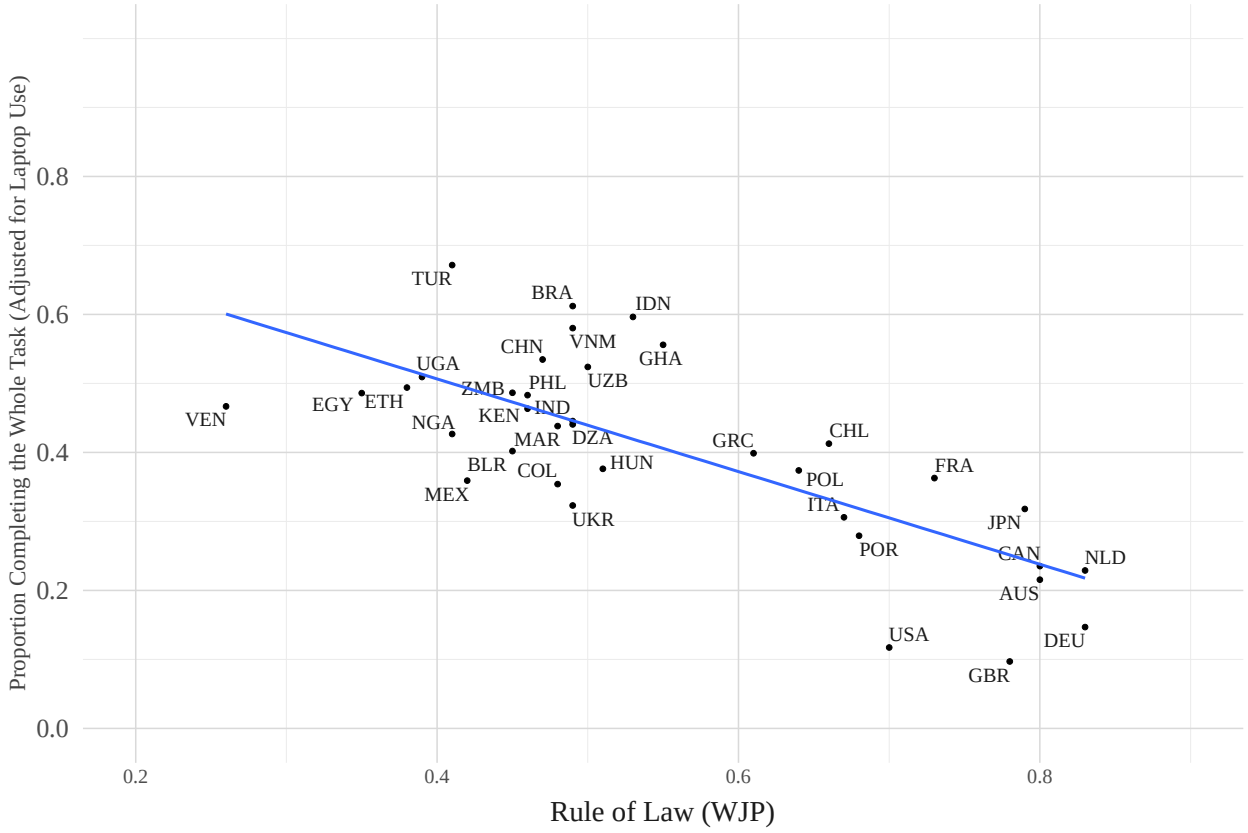
COUNTRY-LEVEL CORRELATIONS IN STUDY 5A

Fig. S15a: **Study 5A, Correlation Between Responsibilism and Marginal Completion Rates (Adjusted for Device Type).**



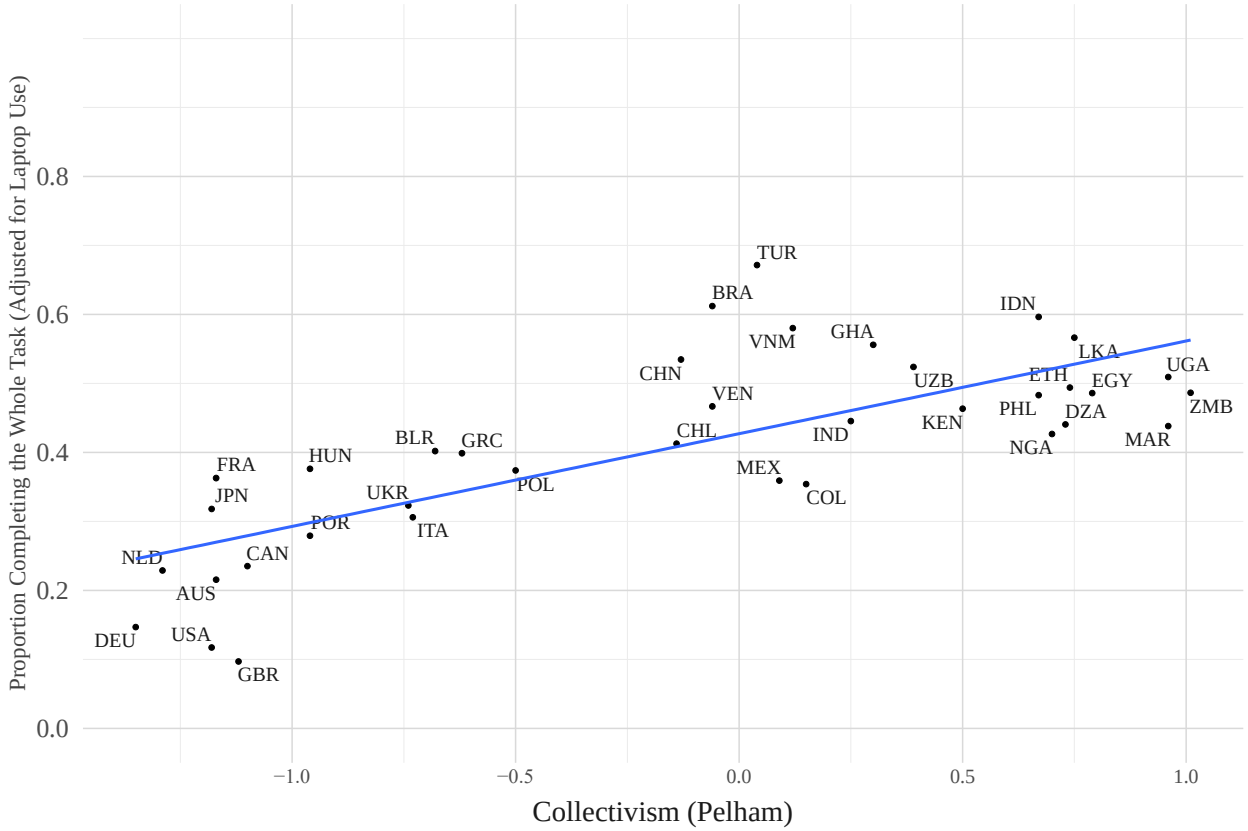
Correlation between marginal completion rates of the data entry task in Study 5A, grouped by country, and responsibilism taken from Talhelm and colleagues (Talhelm et al., under review).

Fig. S15b: **Study 5A, Correlation Between Rule of Law and Marginal Completion Rates (Adjusted for Device Type).**



Correlation between marginal completion rates of the data entry task in Study 5A, grouped by country, and the rule of law index taken from World Justice Project data (World Justice Project, 2023).

***Fig. S15c*: Study 5A, Correlation Between Collectivism and Marginal Completion Rates (Adjusted for Device Type).**



Correlation between marginal completion rates of the data entry task in Study 5A, grouped by country, and the country-level collectivism scores taken from Pelham and colleagues (Pelham et al., 2022).

REFERENCES

Adams, B. G., Van de Vijver, F. J., & De Bruin, G. P. (2012). Identity in South Africa: Examining self-descriptions across ethnic groups. *International Journal of Intercultural Relations*, *36*(3), 377–388.

Adler, N. J., & Jelinek, M. (1986). Is "organization culture" culture bound? *Human Resource Management*, *25*(1), 73–90.

Akcigit, U., Alp, H., & Peters, M. (2021). Lack of selection and limits to delegation: Firm dynamics in developing countries. *American Economic Review*, *111*(1), 231–275.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, *97*(4), 543–569.

Alesina, A., & Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, *53*(4), 898–944.

Altbach, P. G. (2009). The Giants Awake: Higher Education Systems in China and India. *Economic and Political Weekly*, *44*(23), 39–51.

Appelbaum, E., & Schmitt, J. (2009). Review article: Low-wage work in high-income countries: Labor-market institutions and business strategy in the US and Europe. *Human Relations*, *62*(12), 1907–1934. https://doi.org/10.1177/0018726709349200

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization*, *67*(3–4), 671–677.

Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, *120*, 1–17.

Auden, W. H. (1941). *The Dyer's Hand and Other Essays*. Random House.

Baktash, M. B., Heywood, J. S., & Jirjahn, U. (2022). Worker stress and performance pay: German survey evidence. *Journal of Economic Behavior & Organization*, *201*, 276–291. https://doi.org/10.1016/j.jebo.2022.07.003

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, *77*(2), 417–458.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., & Sulaiman, M. (2017). Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, *132*(2), 811–870.

Bandiera, O., Dahlstrand-Rudin, A., & Fischer, G. (2020). Incentives and Culture: Evidence from a Multi-Country Field Experiment. *Working Paper*. https://doi.org/10.1257/rct.4685-2.1

Bandiera, O., Kotia, A., Lindenlaub, I., Moser, C., & Prat, A. (2024). *Meritocracy across Countries*. National Bureau of Economic Research.

Baranowska-Rataj, A., Elekes, Z., & Eriksson, R. (2023). Escaping from low-wage employment: The role of co-worker networks. *Research in Social Stratification and Mobility*, *83*, 100747.

Barnes, B. R., Yen, D., & Zhou, L. (2011). Investigating guanxi dimensions and relationship outcomes: Insights from Sino-Anglo business relationships. *Industrial Marketing Management*, *40*(4), 510–521. https://doi.org/10.1016/j.indmarman.2010.12.007

Baron, J. N. (1988). The employment relation as a social relation. *Special Conference Issue Saving: Its Determinants and Macroeconomic Implications (Part 2 of 2 Parts)*, *2*(4), 492–525. https://doi.org/10.1016/0889-1583(88)90004-4

Baron, J. N. (2013). Empathy wages?: Gratitude and gift exchange in employment relationships. *Research in Organizational Behavior*, *33*, 113–134.

Baron, J. N., & Kreps, D. M. (2012). Employment as an economic and a social relationship. In R. Gibbons & J. Roberts (Eds.), *Handbook of Organizational Economics* (pp. 315–341). Princeton University Press.

Barron, K., & Chou, S. Y. (2016). Developing normative commitment as a consequence of receiving help-the moderated mediating roles of team-member exchange and individualism? Collectivism: A multi-level model. *Journal of Management Sciences*, *3*(2), 141–158.

Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, *69*(1), 25–39.

Bass, B. M. (1997). Does the transactional–transformational leadership paradigm transcend organizational and national boundaries? *American Psychologist*, *52*(2), 130.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.

Benet-Martínez, V., Leu, J., Lee, F., & Morris, M. W. (2002). Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, *33*(5), 492–516.

Berg, J., Furrer, M., Harmon, E., Rani, U., & Silberman, M. S. (2018). *Digital labour platforms and the future of work.* International Labour Office. http://wtf.tw/text/digital_labour_platforms_and_the_future_of_work.pdf

Bertrand, M., & Schoar, A. (2006). The Role of Family in Family Firms. *Journal of Economic Perspectives*, *20*(2), 73–96.

Besley, T., & Ghatak, M. (2008). Status Incentives. *American Economic Review*, *98*(2), 206–211. https://doi.org/10.1257/aer.98.2.206

Bible, New Living Translation. (1996). *Ecclesiastes*. Tyndale House Publishers.

Biernacki, R. (1995). *The fabrication of labor: Germany and Britain, 1640-1914* (Vol. 22). Univ of California Press.

Binmore, K. (2005). Economic man–or straw man? *Behavioral and Brain Sciences*, *28*(6), 817–818.

Blau, P. M., & Scott, W. R. (1962). *Formal organizations: A comparative approach.* Chandler.

Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C. C., Udry, C., & Vaillant, J. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, *605*(7909), 291–297.

Bouchrika, I., Harrati, N., Wanick, V., & Wills, G. (2021). Exploring the impact of gamification on student engagement and involvement with e-learning systems. *Interactive Learning Environments*, *29*(8), 1244–1257.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, *36*(1), 75–111.

Boyacigiller, N. A., & Adler, N. J. (1991). The parochial dinosaur: Organizational science in a global context. *Academy of Management Review*, *16*(2), 262–290.

Boyer, P., & Petersen, M. B. (2018). Folk-economic beliefs: An evolutionary cognitive model. *Behavioral and Brain Sciences*, *41*, 1–51.

Brady, L. M., Fryberg, S. A., & Shoda, Y. (2018). Expanding the interpretive power of psychological science by attending to culture. *Proceedings of the National Academy of Sciences*, *115*(45), 11406–11413.

Bubonya, M., Cobb-Clark, D. A., & Wooden, M. (2017). Mental health and productivity at work: Does what you do matter? *Labour Economics*, *46*, 150–165.

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., & Simon, S. (2022). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, *149*(3–4), 174–205.

Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy*, *127*(4), 1641–1683.

Butcher, K. F., & Schanzenbach, D. W. (2018). Most workers in low-wage labor market work substantial hours, in volatile jobs. *Washington, DC: Center on Budget and Policy Priorities*.

California Department of Industrial Relations. (2021). *Garment Worker Protection Act*. https://www.dir.ca.gov/dlse/Garment/

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., & Wengström, E. (2021). Monetary incentives increase COVID-19 vaccinations. *Science*, *374*(6569), 879–882.

Caria, S. A., & Falco, P. (2022). Skeptical employers: Experimental evidence on biased beliefs constraining firm growth. *Review of Economics and Statistics*, 1–45.

Cassar, L., & Meier, S. (2018). Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives*, *32*(3), 215–238.

Chandrasekhar, A. G., Morten, M., & Peter, A. (2020). *Network-based hiring: Local benefits; global costs*. National Bureau of Economic Research.

Chatman, J. A., & Barsade, S. G. (1995). Personality, organizational culture, and cooperation: Evidence from a business simulation. *Administrative Science Quarterly*, *40*(3), 423–443.

Chen, H., Bolton, L. E., Ng, S., Lee, D., & Wang, D. (2018). Culture, relationship norms, and dual entitlement. *Journal of Consumer Research*, *45*(1), 1–20.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, *103*(2), 690–731.

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*, *11*(7), 928–937.

Chiang, F. F., & Birtch, T. A. (2012). The performance implications of financial and non-financial rewards: An Asian Nordic comparison. *Journal of Management Studies*, *49*(3), 538–570.

Corney, W. J., & Richards, C. H. (2005). A comparative analysis of the desirability of work characteristics: Chile versus the United States. *International Journal of Management*, *22*(2), 159–165.

Dalton, P. S., Rüschenpöhler, J., Uras, B., & Zia, B. (2021). Curating local knowledge: Experimental evidence from small retailers in Indonesia. *Journal of the European Economic Association*, *19*(5), 2622–2657.

Daniels, A. K. (1987). Invisible work. *Social Problems*, *34*(5), 403–415.

Davies, E., & Fafchamps, M. (2022). *Material incentives and effort choice: Evidence from an online experiment across countries*. National Bureau of Economic Research.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 19–43.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, *85*(2), 1029–1069.

DeVoe, S. E., & Iyengar, S. S. (2004). Managers' theories of subordinates: A cross-cultural examination of manager perceptions of motivation and appraisal of performance. *Organizational Behavior and Human Decision Processes*, *93*(1), 47–61.

DiMaggio, P. (1997). Culture and Cognition. *Annual Review of Sociology*, *23*, 263–287.

Domenech Rodríguez, M. M., Baumann, A. A., Vázquez, A. L., Amador-Buenabad, N. G., Franceschi Rivera, N., Ortiz-Pons, N., & Parra-Cardona, J. R. (2018). Scaling out evidence-based interventions outside the US mainland: Social justice or Trojan horse? *Journal of Latina/o Psychology*, *6*(4), 329–344.

Durkheim, É. (1984). *The Division of Labor in Society*. Free Press.

Eaton, L., & Louw, J. (2000). Culture and self in South Africa: Individualism-collectivism predictions. *The Journal of Social Psychology*, *140*(2), 210–217.

Einolf, C. J. (2017). Cross-national differences in charitable giving in the west and the world. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *28*, 472–491.

Erkal, N., Gangadharan, L., & Koh, B. H. (2018). Monetary and non-monetary incentives in real-effort tournaments. *European Economic Review*, *101*, 528–545.

Estrada-Villalta, S., & Adams, G. (2018). Decolonizing development: A decolonial approach to the psychology of economic inequality. *Translational Issues in Psychological Science*, *4*(2), 198–209.

Etkin, J., & Memmi, S. A. (2021). Goal conflict encourages work and discourages leisure. *Journal of Consumer Research*, *47*(5), 716–736.

Falk, A., Gächter, S., & Kovács, J. (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, *20*(3), 251–284.

Fernández, R., & Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, *1*(1), 146–177.

Ferraro, F., Pfeffer, J., & Sutton, R. I. (2005). Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management Review*, *30*(1), 8–24.

Fishbach, A., & Choi, J. (2012). When thinking about goals undermines goal pursuit. *Organizational Behavior and Human Decision Processes*, *118*(2), 99–107.

Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing.* Free Press.

Fisman, R., & Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy*, *115*(6), 1020–1048.

Franklin, B. (1748). *Advice to a Young Tradesman, Written by an Old One*.

Freeman, R. B. (2010). Labor regulations, unions, and social protection in developing countries: Market distortions or efficient institutions? *Handbook of Development Economics*, *5*, 4657–4702.

Frey, B. S., & Gallus, J. (2017). Towards an economics of awards. *Journal of Economic Surveys*, *31*(1), 190–200.

Friedmann, G. (1960). *Qu'est-ce que le travail? 15*(4), 684–701.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, *126*(4), 1755–1798.

Fuchs, C., & Sevignani, S. (2013). What is digital labour? What is digital work? What's their difference? And why do these questions matter for understanding social media? *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *11*(2), 237–293.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Free Press.

Gallus, J. (2017). Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Management Science*, *63*(12), 3999–4015.

Gallus, J., Reiff, J., Kamenica, E., & Fiske, A. P. (2022). Relational incentives theory. *Psychological Review*, *129*(3), 586–602.

Gardner, D. G., Huang, G.-H., Niu, X., Pierce, J. L., & Lee, C. (2015). Organization-based self-esteem, psychological contract fulfillment, and perceived employment opportunities: A test of self-regulatory theory. *Human Resource Management*, *54*(6), 933–953.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., & Arnadottir, J. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104.

Ghani, T., & Reed, T. (2022). Relationships on the rocks: Contract evolution in a market for ice. *American Economic Journal: Microeconomics*, *14*(1), 330–365.

Glynn, M. A. (1994). Effects of work task cues and play task cues on information processing, judgment, and motivation. *Journal of Applied Psychology*, *79*(1), 34–45.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308.

Godelier, M. (1980). *Work and its representations: A research proposal*. *10*(1), 164–174.

Gorodnichenko, Y., & Roland, G. (2011). Individualism, innovation, and long-run growth. *Proceedings of the National Academy of Sciences*, *108*(supplement_4), 21316–21319.

Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, *8*, 125–130.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *The American Economic Review*, *83*(3), 525–548.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, *20*(2), 23–48.

Hansen, F., Smith, M., & Hansen, R. B. (2002). Rewards and recognition in employee motivation. *Compensation & Benefits Review*, *34*(5), 64–72.

Harguess, J. M., Crespo, N. C., & Hong, M. Y. (2020). Strategies to reduce meat consumption: A systematic literature review of experimental studies. *Appetite*, *144*, 104478.

Harwick, C. (2023). Money's mutation of the modern moral mind: The Simmel hypothesis and the cultural evolution of WEIRDness. *Journal of Evolutionary Economics*, *33*(5), 1571–1592. https://doi.org/10.1007/s00191-023-00844-4

Heath, C. (1999). On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes*, *78*(1), 25–62.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., & Ensminger, J. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*(6), 795–815.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). *The Motivation to Work*. John Wiley & Sons.

Hitokoto, H. (2016). Indebtedness in cultural context: The role of culture in the felt obligation to reciprocate. *Asian Journal of Social Psychology, 19*(1), 16–25.

Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorization processes in organizational contexts. *The Academy of Management Review*, *25*(1), 121–140. https://doi.org/10.2307/259266

Hong, Y., Morris, M. W., Chiu, C., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, *55*(7), 709–720.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage Publications.

Hui, C., Lee, C., & Rousseau, D. M. (2004). Psychological contract and organizational citizenship behavior in China: Investigating generalizability and instrumentality. *Journal of Applied Psychology, 89*(2), 311–321.

Hur, J. D., Lee-Yoon, A., & Whillans, A. V. (2021). Are they useful? The effects of performance incentives on the prioritization of work versus personal ties. *Organizational Behavior and Human Decision Processes*, *165*, 103–114.

Hutson, D. (2000). New incentives are on the rise. *Compensation & Benefits Review*, *32*(5), 40–46.

International Labour Organization. (2021a). *Agriculture; Plantations; Other Rural Sectors*. https://www.ilo.org/industries-and-sectors/agriculture-plantations-other-rural-sectors

International Labour Organization. (2021b). *ILO Global Estimates on International Migrant Workers: Results and Methodology*. International Labour Organization. https://www.ilo.org/global/publications/books/WCMS_808935/lang--en/index.htm

International Monetary Fund. (2023). *World Economic Outlook Database*. International Monetary Fund. https://www.imf.org/en/Publications/WEO/weo-database/2023

Ipeirotis, P. G. (2010). Demographics of Mechanical Turk. *Microeconomics: Search; Learning; Information Costs & Specific Knowledge; Expectation & Speculation eJournal*.

Jackson, J. C., Gelfand, M., & Ember, C. R. (2020). A global analysis of cultural tightness in non-industrial societies. *Proceedings of the Royal Society B*, *287*(1930), 20201036.

Jenkins Jr, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, *83*(5), 777–787.

Jevons, W. S. (1871). *The Theory of Political Economy*. McMaster University Archive for the History of Economic Thought. https://ideas.repec.org/b/hay/hetboo/jevons1871.html

Jia, F., & Zsidisin, G. A. (2014). Supply relational risk: What role does guanxi play? *Journal of Business Logistics*, *35*(3), 259–267.

Jiang, L., Wagner, C., & Nardi, B. (2015). Not just in it for the money: A qualitative investigation of workers' perceived benefits of micro-task crowdsourcing. *2015 48th Hawaii International Conference on System Sciences*, 773–782. https://doi.org/10.1109/HICSS.2015.98

Johnson, N. B., & Droege, S. (2004). Reflections on the generalization of agency theory: Cross-cultural considerations. *Human Resource Management Review*, *14*(3), 325–335.

Jose and Roxanne. (2010). *Flowers* [Graphic]. https://www.flickr.com/photos/nyc_xmas/4715530598/

Jung, D. I., & Avolio, B. J. (1999). Effects of leadership style and followers' cultural orientation on performance in group and individual task conditions. *Academy of Management Journal*, *42*(2), 208–218.

Kapoor, H., Rezaei, S., Gurjar, S., Tagat, A., George, D., Budhwar, Y., & Puthillam, A. (2023). Does incentivization promote sharing "true" content online? *Harvard Kennedy School Misinformation Review*, *4*(4). https://doi.org/10.37016/mr-2020-120

Karlsson, N., Loewenstein, G., & McCafferty, J. (2004). The economics of meaning. *Nordic Journal of Political Economy*, *30*(1), 61–75.

Kasser, T., Cohn, S., Kanner, A. D., & Ryan, R. M. (2007). Some costs of American corporate capitalism: A psychological exploration of value and goal conflicts. *Psychological Inquiry*, *18*(1), 1–22.

King, R. C., & Bu, N. (2005). Perceptions of the mutual obligations between employees and employers: A comparative study of new generation IT professionals in China and the United States. *The International Journal of Human Resource Management*, *16*(1), 46–64.

Kizilcec, R. F., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences*, *114*(17), 4348–4353.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191–210.

Komlosy, A., Watson, J. K., & Balhorn, L. (2018). *Work: The Last 1,000 Years*. Verso Books. http://ebookcentral.proquest.com/lib/uchicago/detail.action?docID=6048740

Kosfeld, M., Neckermann, S., & Yang, X. (2017). The effects of financial and recognition incentives across work contexts: The role of meaning. *Economic Inquiry*, *55*(1), 237–247.

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949.

Kreps, D. M. (1997). Intrinsic Motivation and Extrinsic Incentives. *The American Economic Review*, *87*(2), 359–364.

Laran, J., & Janiszewski, C. (2011). Work or fun? How task construal and completion influence regulatory behavior. *Journal of Consumer Research*, *37*(6), 967–983.

Leary, M. R., Jongman-Sereno, K. P., & Diebels, K. J. (2014). The pursuit of status: A self-presentational perspective on the quest for social value. *The Psychology of Social Status*, 159–178.

Ledeneva, A. (1998). *Russia's Economy of Favours: Blat, Networking and Informal Exchange*. Cambridge University Press. https://books.google.com/books?id=ENMhO5aCkLAC

Ledeneva, A. (2008). " Blat" and" Guanxi": Informal Practices in Russia and China. *Comparative Studies in Society and History*, *50*(1), 118–144.

Li, Y., & Zahra, S. A. (2012). Formal institutions, culture, and venture capital activity: A cross-country analysis. *Journal of Business Venturing*, *27*(1), 95–111.

Liebe, U., Gewinner, J., & Diekmann, A. (2018). What is missing in research on non-monetary incentives in the household energy sector? *Energy Policy*, *123*, 180–183.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, *57*(3), 426–441.

Lukács, G., 1885-1971. (1971). *History and class consciousness; studies in Marxist dialectics*. Cambridge, Mass. : MIT Press, [1971]. https://search.library.wisc.edu/catalog/999467346302121

Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One individual, two identities: Frame switching among biculturals. *Journal of Consumer Research*, *35*(2), 279–293.

Macchiavello, R. (2022). Relational contracts and development. *Annual Review of Economics*, *14*, 337–362.

Madan, S., Ma, A., Pandey, N., Rattan, A., & Savani, K. (2023). Support for increasing low-wage workers' compensation: The role of fixed-growth mindsets about intelligence. *Journal of Experimental Psychology: General*, *152*(4), 935–955. https://doi.org/10.1037/xge0001303

Mandel, D. R. (2006). Economic transactions among friends: Asymmetric generosity but not agreement in buyers' and sellers' offers. *Journal of Conflict Resolution*, *50*(4), 584–606.

Markus, H. R. (2016). What moves people to action? Culture and motivation. *Current Opinion in Psychology*, *8*, 161–166.

Marx, K. (1867). *Capital: A Critique of Political Economy*. Verlag von Otto Meisner.

MasterPhoto-DK. (2007). *Building 1* [Graphic]. https://www.flickr.com/photos/euromagic/436810411/

Mauss, M. (2000). *The gift: The form and reason for exchange in archaic societies*. WW Norton & Company.

Mayo, E. (1933). *The Social Problems of an Industrial Civilization*. Macmillan.

Medvedev, D., Davenport, D., Talhelm, T., & Li, Y. (2024). The motivating effect of monetary over psychological incentives is stronger in WEIRD cultures. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01769-5

Meeker, B. F. (1971). Decisions and exchange. *American Sociological Review, 36*(3), 485–495.

Merriam-Webster. (2024). *Work*. https://www.merriam-webster.com/dictionary/work

Michailova, S., & Worm, V. (2003). Personal Networking in Russia and China: Blat and Guanxi. *European Management Journal, 21*(4), 509–519.

Miles, A. (2015). The (re) genesis of values: Examining the importance of values for action. *American Sociological Review, 80*(4), 680–704.

Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., & Beshears, J. (2021). Megastudies improve the impact of applied behavioural science. *Nature, 600*(7889), 478–483.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54*(12), 1053–1060.

Minkov, M., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., Khassenbekov, Y., & Mudd, B. (2017). A revision of Hofstede's individualism-collectivism dimension: A new national index from a 56-country study. *Cross Cultural & Strategic Management, 24*(3), 386–404.

Morris, M. W., Hong, Y., Chiu, C., & Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*, *129*, 1–13.

Morschheuser, B., & Hamari, J. (2019). The gamification of work: Lessons from crowdsourcing. *Journal of Management Inquiry*, *28*(2), 145–148.

Moyce, S. C., & Schenker, M. (2018). Migrant Workers and Their Occupational Health and Safety. In *Annual Review of Public Health* (Vol. 39, Issue Volume 39, 2018, pp. 351–365). Annual Reviews. https://doi.org/10.1146/annurev-publhealth-040617-013714

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, *31*(6), 678–701.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38.

OECD. (2022). *Purchasing power parities (PPP): Total, National currency units/US dollar, 2000 – 2022*. [Dataset]. OECD. https://doi.org/10.1787/1290ee5a-en

Oyserman, D. (2015). Values, psychology of. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., Vol. 25, pp. 36–40). Elsevier.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking Individualism and Collectivism: Evaluation of Theoretical Assumptions and Meta-Analyses. *Psychological Bulletin*, *128*(1), 3–72.

Pelham, B., Hardin, C., Murray, D., Shimizu, M., & Vandello, J. (2022). A truly global, non-WEIRD examination of collectivism: The global collectivism index (GCI). *Current Research in Ecological and Social Psychology*, *3*, 100030.

Polanyi, K. (1957a). The economy as instituted process. In K. Polanyi, C. M. Arensberg, & H. W. Pearson (Eds.), *Trade market in the Early Empires* (pp. 243–270). The Free Press.

Polanyi, K. (1957b). *The Great Transformation: The Political and Economic Origins of Our Time*. Beacon Press. https://books.google.com/books?id=YfpIs1Z6B2sC

Prelec, D., & Bodner, R. (2003). Self-signaling and self-control. In *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice* (pp. 277–298).

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401–11405.

Raja, U., Johns, G., & Ntalianis, F. (2004). The impact of personality on psychological contracts. *Academy of Management Journal*, *47*(3), 350–367.

Ralston, D. A., Cunniff, M. K., & Gustafson, D. J. (1995). Cultural accommodation: The effect of language on the responses of bilingual Hong Kong Chinese managers. *Journal of Cross-Cultural Psychology*, *26*(6), 714–727.

Rama, M. G. (2003). Globalization and workers in developing countries. *Policy Research Working Paper Series*, *2958*.

Rani, U., & Furrer, M. (2021). Digital labour platforms and new forms of flexible work in developing countries: Algorithmic management of work and workers. *Competition & Change*, *25*(2), 212–236.

Rodríguez, C. M., & Wilson, D. T. (2002). Relationship bonding and trust as a foundation for commitment in US–Mexican strategic alliances: A structural equation modeling approach. *Journal of International Marketing*, *10*(4), 53–76.

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*, 321–328. https://doi.org/10.1609/icwsm.v5i1.14105

Ross, M., & Bateman, N. (2019). *Meet the low-wage workforce*. Brookings Institution. https://www.brookings.edu/research/meet-the-low-wage-workforce/

Rosso, B. D., Dekas, K. H., & Wrzesniewski, A. (2010). On the meaning of work: A theoretical integration and review. *Research in Organizational Behavior*, *30*, 91–127.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *The American Economic Review*, *51*(8), 1068–1095.

Rousseau, D. M., & McLean Parks, J. (1993). The contracts of individuals and organizations. *Research in Organizational Behavior*, *15*, 1–43.

Schulz, J., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018). The origins of WEIRD psychology. *Available at SSRN 3201031*.

Scott, J. C. (1976). *The moral economy of the peasant: Rebellion and subsistence in Southeast Asia*. Yale University Press. http://www.jstor.org/stable/j.ctt1bh4cdk

Shao, L., Kwok, C. C., & Guedhami, O. (2010). National culture and dividend policy. *Journal of International Business Studies*, *41*, 1391–1414.

Shen, H., Wan, F., & Wyer Jr, R. S. (2011). Cross-cultural differences in the refusal to accept a small gift: The differential influence of reciprocity norms on Asians and North Americans. *Journal of Personality and Social Psychology*, *100*(2), 271.

Si, A. (2011). A diachronic investigation of Hindi–English code-switching, using Bollywood film scripts. *International Journal of Bilingualism*, *15*(4), 388–407.

Simmel, G. (1907). *The Philosophy of Money* (T. Bottomore & D. Frisby, Trans.; Second Enlarged Edition). Routledge.

Singelis, T. M. (1994). The Measurement of Independent and Interdependent Self-Construals. *Personality and Social Psychology Bulletin*, *20*(5), 580–591.

Sinha, J. B. P. (1984). Towards partnership for relevant research in the third world. *International Journal of Psychology*, *19*(1–4), 169–177.

Sloan, T. S. (1990). Psychology for the third world? *Journal of Social Issues*, *46*(3), 1–20.

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, *74*(2), 482–493.

Takano, Y., & Sogon, S. (2008). Are Japanese more collectivistic than Americans? Examining conformity in in-groups and the reference-group effect. *Journal of Cross-Cultural Psychology*, *39*(3), 237–250.

Talhelm, T., Wei, L., Sun, R., Medvedev, D., San Martin, A., Samekin, A., Scherman, A. Z., & Responsibiism Study Team. (under review). *Collectivism Isn't What People Think It Is: A Study of 100 Cultures*.

Tang, T. L., & Baumeister, R. F. (1984). Effects of personal values, perceived surveillance, and task labels on task preference: The ideology of turning play into work. *Journal of Applied Psychology*, *69*(1), 99–105.

Thaler, R. H. (2000). From homo economicus to homo sapiens. *Journal of Economic Perspectives*, *14*(1), 133–141.

Thibault Landry, A., Schweyer, A., & Whillans, A. (2017). Winning the war for talent: Modern motivational methods for attracting and retaining employees. *Compensation & Benefits Review*, *49*(4), 230–246.

Thomas, C. C., & Markus, H. R. (2023). Enculturating the science of international development: Beyond the WEIRD independent paradigm. *Journal of Cross-Cultural Psychology*, *54*(2), 195–214.

Thomas, D. C., Au, K., & Ravlin, E. C. (2003a). Cultural variation and the psychological contract. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(5), 451–471.

Thomas, D. C., Au, K., & Ravlin, E. C. (2003b). Cultural variation and the psychological contract. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(5), 451–471.

Thompson, E. P. (1971). The moral economy of the English crowd in the eighteenth century. *Past & Present*, *50*, 76–136.

Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., Becker, J. C., Becker, M., Chiu, C., & Choi, H.-S. (2018). Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proceedings of the National Academy of Sciences*, *115*(29), 7521–7526.

Tiessen, J. H. (1997). Individualism, collectivism, and entrepreneurship: A framework for international comparative research. *Journal of Business Venturing*, *12*(5), 367–384.

Touré-Tillery, M., & Fishbach, A. (2018). Three sources of motivation. *Consumer Psychology Review*, *1*(1), 123–134.

Trafimow, D., Silverman, E. S., Fan, R. M.-T., & Fun Law, J. S. (1997). The effects of language and priming on the relative accessibility of the private self and the collective self. *Journal of Cross-Cultural Psychology*, *28*(1), 107–123.

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology*, *54*(2), 323–338.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*(1), 118–128.

Vasmer, M. (1950). *Etymological Dictionary of the Russian Language* (Vols. 1–4). Progress.

Verkuyten, M., & Pouliasi, K. (2002). Biculturalism among older children: Cultural frame switching, attributions, self-identification, and attitudes. *Journal of Cross-Cultural Psychology*, *33*(6), 596–609.

Vuolevi, J. H., & Van Lange, P. A. (2010). Beyond the information given: The power of a belief in self-interest. *European Journal of Social Psychology*, *40*(1), 26–34.

Walo, S. (2023). Historical Changes in the Meaning of Work: Quantifying Smith, Marx and Weber with Word Embeddings. *Preprints of the Open Science Framework*. https://doi.org/10.31235/osf.io/pxj7m

Walumbwa, F. O., Lawler, J. J., & Avolio, B. J. (2007). Leadership, individual differences, and work-related attitudes: A cross-culture investigation. *Applied Psychology*, *56*(2), 212–230.

Weber, M. (1905). *The Protestant ethic and the spirit of capitalism*. Routledge.

Westwood, R., Chan, A., & Linstead, S. (2004). Theorizing Chinese employment relations comparatively: Exchange, reciprocity and the moral economy. *Asia Pacific Journal of Management*, *21*, 365–389.

White, K., & Lehman, D. R. (2005). Culture and social comparison seeking: The role of self-motives. *Personality and Social Psychology Bulletin*, *31*(2), 232–242.

Wiepking, P. (2021). The global study of philanthropic behavior. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *32*(2), 194–203.

Wilmers, N., & Zhang, L. (2022). Values and inequality: Prosocial jobs and the college wage premium. *American Sociological Review*, *87*(3), 415–442.

Woolley, K., & Fishbach, A. (2021). *How the promise of incentives affects inferences about task enjoyment.* [Working Paper].

World Bank. (2020). *Adjusted net national income per capita (current US$)*. https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.CD

World Bank. (2021). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators

World Bank. (2024). *Digital Progress and Trends Report 2023*. The World Bank. doi:10.1596/978-1-4648-2049-6

World Justice Project. (2023). *Rule of Law Index*. World Justice Project. https://worldjusticeproject.org/our-work/research-and-data/wjp-rule-law-index-2023

Wu, S. J., & Paluck, E. L. (2021). Designing nudges for the context: Golden coin decals nudge workplace behavior in China. *Organizational Behavior and Human Decision Processes*, *163*, 43–50.

Zaki, J., Neumann, E., & Baltiansky, D. (2021). Market cognition: How exchange norms alter social experience. *Current Directions in Psychological Science*, *30*(3), 236–241.